University of Alberta

Molecular and Cytogenetic Analysis of the Rat Alpha-2u-Globulin Locus

by

David Anson McFadyen

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of Doctor of Philosophy

in

Molecular Biology and Genetics

Department of Biological Sciences

Edmonton, Alberta

Spring 1997

National Library of Canada
Acquisitions and Bibliographic Services
395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque nationale du Canada
Acquisitions et services bibliographiques
395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

Our file Notre référence

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of his/her thesis by any means and in any form or format, making this thesis available to interested persons.

The author retains ownership of the copyright in his/her thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced with the author's permission.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de sa thèse de quelque manière et sous quelque forme que ce soit pour mettre des exemplaires de cette thèse à la disposition des personnes intéressées.

L'auteur conserve la propriété du droit d'auteur qui protège sa thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-21600-4

Canada

University of Alberta

Library Release Form

Name of Author: David Anson McFadyen

Title of Thesis: Molecular and Cytogenetic Analysis of the Rat Alpha-2u-Globulin Locus

Degree: Doctor of Philosophy

Year this Degree Granted: 1997

Permission is hereby granted to the University of Alberta Library to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly, or scientific research purposes only.

The author reserves all other publication and other rights in association with the copyright in the thesis, and except as hereinbefore provided, neither the thesis nor any substantial portion thereof may be printed or otherwise reproduced in any material form whatever without the author's prior written permission.

David McFadyen

April 18/97

4312 Rae Street
Regina, SK
S4S 3A8

University of Alberta

Faculty of Graduate Studies and Research

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled "Molecular and Cytogenetic Analysis of the Rat Alpha-2u-Globulin Locus" submitted by David Anson McFadyen in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Molecular Biology and Genetics.

Dr. John Locke (supervisor)

Dr. David Nash

Dr. Charlotte Spencer

Dr. Alan Bateson

Dr. Douglas Dickinson

To my wife Karen, for her unselfish support. You are an inspiration.

ABSTRACT

The alpha-2u-globulins are a group of similar proteins, belonging to the lipocalin superfamily of proteins, that are synthesized in a subset of secretory tissues in rats. These proteins are hypothesized to function as carriers of pheromones, and have been shown to bind numerous small hydrophobic molecules. The many alpha-2u-globulin isoforms are encoded by a group of about 20 genes, that despite extensive sequence homology, show diverse expression patterns involving complex hormonal, tissue-specific, and developmental regulation. Although aspects of the regulation of some members of this gene family have been investigated, the organization of the alpha-2u-globulin gene cluster has not been examined. The cloning and mapping of the entire locus is viewed as a requisite step towards understanding the regulation of this gene family. This thesis is an initial characterization of the alpha-2u-globulin locus using molecular and cytogenetic approaches.

Genomic Southern blot and slot blot analysis suggest there are 20-24 alpha-2u-globulin genes in the rat genome. Fluorescence *In Situ* Hybridization (FISH) was used to show that the alpha-2u-globulin genes are clustered at a single site on chromosome 5 (5q22→24). Southern blots of rat genomic DNA separated by pulsed field gel electrophoresis indicated that the

alpha-2u-globulin genes are contained on two Nru I fragments with a total size of 880 kb.

Analysis of three P1 clones containing alpha-2u-globulin genes indicated that the alpha-2u-globulin genes are tandemly arranged. FISH to deproteinized rat genomic DNA fibers showed this was the predominant organizational pattern of the alpha-2u-globulin genes in this cluster. The organization of the alpha-2u-globulin genes in the rat as a tandem array of single genes differs from the homologous major urinary protein genes in the mouse, which are organized as a tandem array of divergently oriented gene pairs.

The contrasting organization of these two loci is consistent with the rapid amplification, or reamplification, from different genes, or gene sets preexisting in the common ancestor of these species. Such rapid changes in the structure of these gene clusters may have consequences for the proposed function of the protein products encoded by these genes.

ACKNOWLEDGMENTS

First and foremost, I am indebted to Dr. John Locke for his willingness to take on my supervision. His support, guidance and countless discussions during the final stages of this project were invaluable. I am grateful for his critical reading of earlier versions of this thesis, and his insightful suggestions which led to numerous improvements.

I want to thank Dr. Bill Addison for introducing me to the complexities of the alpha-2u-globulin gene family, and for his guidance during the early phase of this project.

I greatly appreciate the efforts of the other members of my supervisory committee, Dr. David Nash and Dr. Charlotte Spencer for their helpful contributions and comments at various stages of this project including an earlier version of this thesis.

I would like to express my appreciation to Dr. Jim Lin for allowing me to conduct much of the cytogenetic work reported in this thesis in his laboratory. I must also thank Dr. Charles Lee for introducing me to the intricacies of Fluorescence *In Situ* Hybridization.

Thank you to my fellow graduate students, and friends, Mary Christopher, and Kathy Wang for the many, and wide ranging conversations that helped relieve the stresses of graduate student life. I also want to thank Mark Hicks, Andrew Simmonds, and Jeff Hoyem for their assistance with the assorted computer applications.

I am also grateful for the opportunities provided me by the Departments of Genetics, and Biological Sciences, to impart my enthusiasm for science to others, and develop my instructional skills through the teaching of laboratory and seminar sessions. In this respect, I want to thank Laura Querenguesser for the guidance she provided me on becoming an effective instructor and communicator.

Completing this project would not have been possible without the enduring encouragement and support I received from my family throughout my studies-thank you to **all** of you for that generous gift!

Support from the Natural Sciences and Engineering Research Council of Canada, the Government of Alberta, the Faculty of Graduate Studies and Research University of Alberta, and the Departments of Genetics and Biological Sciences at various stages of these studies is gratefully acknowledged.

TABLE OF CONTENTS

CHAPTER 1. INTRODUCTION ---- 1

1.1. THE EUKARYOTIC GENOME ---- 1
1.1.1. EARLY STUDIES ---- 1
1.1.2. RAPIDLY REANNEALING SEQUENCES ---- 2
1.1.3. INTERMEDIATELY REANNEALING SEQUENCES ---- 2
1.1.4. SLOWLY REANNEALING SEQUENCES ---- 5
1.1.5. MAP CONSTRUCTION ---- 5
1.1.5.1. GENERAL ---- 5
1.1.5.2. LINKAGE MAPS ---- 6
1.1.5.3. PHYSICAL MAPS ---- 6
1.1.5.4. CYTOGENETIC MAPS ---- 9
1.1.6. THE RAT GENETIC MAP ---- 10

1.2. MULTIGENE FAMILIES ---- 11
1.2.1. INTRODUCTION ---- 11
1.2.2. GENOMIC ORGANIZATION OF MULTIGENE FAMILIES ---- 11
1.2.3. EVOLUTION OF MULTIGENE FAMILIES ---- 12
1.2.4. RELATIONSHIP BETWEEN GENE NUMBER AND ORGANIZATION ---- 16
1.2.5. RELATIONSHIP BETWEEN ORGANIZATIONAL ARRANGEMENTS ---- 16

1.3. THE LIPOCALINS ---- 17
1.3.1. STRUCTURE OF LIPOCALIN PROTEINS ---- 17
1.3.2. LIPOCALIN PROTEIN SEQUENCE CONSERVATION ---- 23
1.3.3. LIPOCALIN FUNCTION ---- 25
1.3.4. LIPOCALIN GENOMIC ORGANIZATION ---- 26

1.4. RAT ALPHA-2U-GLOBULIN ---- 28
1.4.1. INTRODUCTION ---- 28
1.4.2. HORMONAL REGULATION OF HEPATIC ALPHA-2U-GLOBULIN ---- 29
1.4.2.1. THYROID HORMONE ---- 29
1.4.2.2. GROWTH HORMONE ---- 30
1.4.2.3. INSULIN ---- 32
1.4.2.4. SEX HORMONES ---- 33
1.4.2.5. GLUCOCORTICOIDS ---- 34
1.4.3. NON-HEPATIC ALPHA-2U-GLOBULIN ---- 37
1.4.4. ALPHA-2U-GLOBULIN AND MUP GENOMIC STRUCTURE ---- 39
1.4.5. ALPHA-2U-GLOBULIN FUNCTION ---- 44

1.5. SCOPE AND NATURE OF THIS WORK ---- 47

CHAPTER 2. MATERIALS AND METHODS ---- 49

2.1. TISSUE CULTURE METHODS ---- 49
2.1.1. CELL LINES ---- 49
2.1.2. GROWTH CONDITIONS ---- 49

2.2. MOLECULAR BIOLOGICAL METHODS ---- 49
2.2.1. RECOMBINANT DNA TECHNIQUES ---- 49
2.2.1.1. BACTERIAL STRAINS ---- 49

2.2.1.2. ALPHA-2U-GLOBULIN CONTAINING PLASMIDS ---- 51
2.2.1.3. PLASMID DNA PREPARATION ---- 51
2.2.1.4. RESTRICTION ENDONUCLEASE DIGESTIONS ---- 54
2.2.1.5. AGAROSE GEL ELECTROPHORESIS ---- 54
2.2.1.6. SOUTHERN BLOTTING ---- 54
2.2.1.7. SLOT BLOTTING ---- 55
2.2.1.8. MEMBRANE PRE-HYBRIDIZATION ---- 55
2.2.1.9. PREPARATION OF RADIO-LABELED PROBE DNA ---- 55
2.2.1.10. HYBRIDIZATION AND WASH PROTOCOLS ---- 56
2.2.1.11. ELECTROPORATION ---- 56
2.2.1.12. IMAGE ANALYSIS PROTOCOL ---- 57
2.2.2. ISOLATION OF GENOMIC DNA ---- 57
2.2.3. RAT LAMBDA GENOMIC LIBRARY ANALYSIS ---- 57
2.2.3.1. ISOLATION OF LAMBDA GENOMIC CLONES CONTAINING ALPHA-2U-GLOBULIN GENE 91 ---- 57
2.2.3.2. DIFFERENTIAL STRINGENCY SCREENING ---- 60
2.2.4. RAT BACTERIOPHAGE P1 LIBRARY ANALYSIS ---- 61
2.2.4.1. DEVELOPMENT OF PCR PRIMERS AND CONDITIONS ---- 61
2.2.4.2. P1 LIBRARY SCREENING ---- 61
2.2.4.3. ISOLATION OF P1 DNA ---- 70
2.2.4.4. P1 CLONE MAPPING ---- 70
2.2.5. GENOMIC CLONE TYPING ---- 71
2.2.5.1. DEVELOPMENT OF PCR PRIMERS AND CONDITIONS ---- 71
2.2.5.2. ANALYSIS OF PCR PRODUCTS ---- 72
2.2.5.3. DENSITOMETRIC ANALYSIS OF PCR SOUTHERN BLOT ---- 72

2.3. PULSED FIELD GEL ANALYSIS OF GENOMIC DNA ---- 72
2.3.1. PREPARATION OF PULSED FIELD GEL BLOCKS ---- 72
2.3.2. RESTRICTION ENDONUCLEASE DIGESTION OF PULSED FIELD AGAROSE BLOCKS ---- 73
2.3.3. ELECTROPHORETIC SEPARATION ---- 73

2.4. FLUORESCENT *IN SITU* HYBRIDIZATION ---- 74
2.4.1. PREPARATION OF METAPHASE SPREADS ---- 74
2.4.2. PREPARATION OF BIOTIN LABELED PROBES ---- 74
2.4.3. *IN SITU* HYBRIDIZATION ---- 75
2.4.4. IMMUNODETECTION ---- 76
2.4.5. IMAGE ACQUISITION ---- 76

2.5. FIBER-FISH ---- 77
2.5.1. PRETREATMENT OF MICROSCOPE SLIDES ---- 77
2.5.2. PREPARATION OF GENOMIC DNA TARGET FIBERS ---- 77
2.5.3. PREPARATION OF P1 DNA TARGET FIBERS ---- 77
2.5.4. PREPARATION OF PROBES ---- 78
2.5.5. *IN SITU* HYBRIDIZATION ---- 78
2.5.6. IMMUNODETECTION ---- 78
2.5.7. IMAGE ACQUISITION ---- 79
2.5.8. IMAGE ANALYSIS ---- 79

CHAPTER 3. RESULTS-MOLECULAR ANALYSES ---- 81

3.1. ESTIMATION OF ALPHA-2U-GLOBULIN GENE NUMBERS ---- 81
3.1.1. SOUTHERN BLOT ANALYSIS ---- 81
3.1.2. SLOT BLOT ANALYSIS ---- 86

3.2. LAMBDA CLONE CHARACTERIZATION ---- 86
3.2.1. HOMOLOGY MAPPING OF LAMBDA GENOMIC SUBCLONES ---- 86
3.2.2. DIFFERENTIAL STRINGENCY SCREENING OF A LAMBDA GENOMIC LIBRARY ---- 92
3.2.3. HOMOLOGOUS UPSTREAM SEQUENCES AMONG ALPHA-2U-GLOBULIN GENES ---- 95
3.2.4. DIVERGENCE IN THE 3´ ENDS OF ALPHA-2U-GLOBULIN GENES ---- 97

3.3. P1 CLONE CHARACTERIZATION ---- 98
3.3.1. SIZE OF P1 INSERTS ---- 98
3.3.2. P1 GENE CONTENT ---- 99
3.3.3. P1 RESTRICTION MAPS ---- 106

3.4. PULSED FIELD GEL MAPPING ---- 106
3.4.1. LONG RANGE MAP OF THE ALPHA-2U-GLOBULIN LOCUS ---- 106
3.4.2. PLACEMENT OF A P1 CLONE ON THE LONG RANGE MAP ---- 111

3.5. ALPHA-2U-GLOBULIN GENE CLASSIFICATION ---- 112
3.5.1. GENOMIC CLONE TYPES ---- 112
3.5.2. PROPORTION OF EACH GENE CLASS IN THE GENOME ---- 115
3.5.3. DISTRIBUTION OF GENE CLASSES IN THE GENOME ---- 120

CHAPTER 4. RESULTS-CYTOGENETIC ANALYSES ---- 124

4.1. RAT METAPHASE FISH ---- 124
4.1.1. LOCALIZATION OF THE ALPHA-2U-GLOBULIN LOCUS ---- 124

4.2. P1 FIBER FISH ---- 124
4.2.1. P1 GENE CONTENT ---- 124
4.2.2. SPACING OF ALPHA-2U-GLOBULIN GENES CARRIED ON P1 CLONES ---- 129

4.3. RAT GENOMIC DNA FIBER FISH ---- 132
4.3.1. ALPHA-2U-GLOBULIN GENE ORIENTATION ---- 132

CHAPTER 5. DISCUSSION ---- 143

5.1. LOCATION OF ALPHA-2U-GLOBULIN AND OTHER LIPOCALIN GENES ---- 143

5.2. HOMOLOGY AMONG ALPHA-2U-GLOBULIN GENES ---- 144
5.2.1. SEQUENCE SIMILARITY ---- 144
5.2.2. CROSS-HYBRIDIZATION OF FLANKING SEQUENCES AMONG ALPHA-2U-GLOBULIN GENES ---- 145

5.3. ORGANIZATION OF THE ALPHA-2U-GLOBULIN GENE CLUSTER ---- 146

5.4. ORIGIN OF THE ALPHA-2U-GLOBULIN REPEAT ---- 150

5.5. EVOLUTIONARY RELATIONSHIP BETWEEN RAT ALPHA-2U-GLOBULIN AND MOUSE MUP GENES ---- 150

5.6. IMPLICATIONS OF ALPHA-2U-GLOBULIN GENE EVOLUTION ON REPRODUCTIVE ISOLATION ---- 153

5.7. FUTURE STUDIES ---------- 156

CHAPTER 6. REFERENCES ---------- 158

LIST OF TABLES

Table 1.1. Examples of Multigene Familes and Their Organization Organization----13
Table 1.2. Representatives of the Lipocalin Superfamily of Proteins----19
Table 2.1. Bacterial Strains----50
Table 2.2. PCR Primer Parameters----62
Table 3.1. Densitometric Analysis of an Alpha-2u-Globulin Quantitative Slot Blot----90
Table 3.2. Analysis of a Rat Genomic Library----96
Table 3.3. Gene Type Expression Patterns----116
Table 3.4. Densitometric Analysis of Restriction Enzyme Digested PCR Products: Determination of Gene Class Proportions--119
Table 4.1. Alpha-2u-Globulin Gene Orientation in the Rat Genome----137
Table A.1. Physical Analysis of Fiber FISH Images of p2860----197
Table A.2. Physical Analysis of Fiber FISH Images of p2861----199
Table A.3. Physical Analysis of Fiber FISH Images of p2862----201

LIST OF FIGURES

Figure 1.1. The lipocalin structure---22
Figure 1.2. Phylogenetic tree of 25 members of the lipocalin family---24
Figure 1.3. Structure of the alpha-2u-globulin gene---41
Figure 1.4. Structure of the MUP 45 kb palindromic unit---43
Figure 2.1. Alpha-2u-globulin containing plasmids---53
Figure 2.2. Restriction maps of 5 subclones from the rat genomic lambda library---59
Figure 2.3. Exonic sequences of 12 alpha-2u-globulin genes reported in Genbank---64
Figure 3.1. Southern hybridization analysis of rat genomic DNA---83
Figure 3.2. Analysis for divergence in the 3′ ends of alpha-2u-globulin genes---85
Figure 3.3. Analysis of the 5′ upstream region of alpha-2u-globulin genes---88
Figure 3.4. Alpha-2u-globulin quantitative slot blot---89
Figure 3.5. Graphical results from an alpha-2u-globulin quantitative slot blot---91
Figure 3.6. Genomic subclone homology mapping---94
Figure 3.7. Analysis of P1 clones---101
Figure 3.8. P1 clone alpha-2u-globulin gene content---103
Figure 3.9. Restriction site polymorphism in the upstream region of alpha-2u-globulin genes---105
Figure 3.10. Restriction maps of three alpha-2u-globulin containing bacteriophage P1 clones---108
Figure 3.11. Southern hybridization analysis of rat genomic DNA separated by pulsed field gel electrophoresis---110
Figure 3.12. Restriction analysis of PCR products amplified from various templates with the primer pair KAW1/KAW2---114
Figure 3.13. Southern blot analysis of PCR products---118
Figure 3.14. Southern analysis of PCR products from pulsed field gel purified templates---122
Figure 4.1. Rat metaphase FISH---126
Figure 4.2. P1 Fiber FISH---128
Figure 4.3. Schematic representation of Fiber FISH to nine p2860 molecules---131
Figure 4.4. Schematic representation of Fiber FISH to ten p2861 molecules---134
Figure 4.5. Schematic representation of Fiber FISH to eleven p2862 molecules---136
Figure 4.6. Possible arrangement of the genes at the alpha-2u-globulin locus---140
Figure 4.7. Six rat genomic Fiber FISH profiles covering the alpha-2u-globulin locus---142

LIST OF SYMBOLS AND ABBREVIATIONS

A	adenine
Asn	asparagine
Asp	aspartic acid
BAC	bacterial artificial chromosome
BAAD	biotinylated anti-avidin
bp	base pair
C	cytosine
CAT	chloramphenicol acetyl transferase
cDNA	complementary deoxyribonucleic acid
CHEF	contour clamped homogenous electric field
Ci	curies
cM	centimorgan
cpm	counts per minute
DAPI	4',6-Diamidine-2'-Phenyl-indole Dihydrochloride
dATP	deoxyadenosine 5'-triphosphate
dCTP	deoxycytidine 5'-triphosphate
dGTP	deoxyguanosine 5'-triphosphate
DMEM	Dulbeccos modified Eagles medium
DNA	deoxyribonucleic acid
Dnase	deoxyribonuclease
dNTP	deoxyribonucleotide 5'-triphosphate
dUTP	deoxyuridine 5'-triphosphate
EDTA	ethylene-diamine-tetraacetic acid
FAD	fluorescene avidin
FISH	fluorescence *in situ* hybridization
G	guanine
Gly	glycine
IPTG	isopropylthio-β-D-galactoside
kb	kilobases
kV	kilovolts
LB	Luria broth
LINE	long interspersed repeated DNA
M	molar
Mb	megabases
mg	milligrams
ml	milliliters
mM	millimolar
mRNA	messenger ribonucleic acid
MUP	major urinary protein
ng	nanograms
OD	optical density
PAC	bacteriophage P1 artificial chromosome
PBS	phophate buffered saline

LIST OF SYMBOLS AND ABBREVIATIONS CONTINUED

PCR	polymerase chain reaction
PI	propidium iodide
pM	picomolar
PMSF	phenyl methylsulfonyl fluoride
RFLP	restriction fragment length polymorphism
RNA	ribonucleic acid
rRNA	ribosomal ribonucleic acid
rpm	revolutions per minute
RTPCR	reverse transcription polymerase chain reaction
SDS	sodium dodecylsulfate
SINE	short interspersed repeated DNA
SSC	standard saline citrate
T	thymine
TB	terrific broth
Thr	threonine
Tris	tris-(hydroxymethyl)-aminomethane
Trp	tryptophan
Tyr	tyrosine
UV	ultra-violet
YAC	yeast artificial chromosome
µCi	microcuries
µg	micrograms
µl	microliters
µm	micrometers
µM	micromolar

LIST OF SUPPLIERS

Adobe	Adobe Systems Inc.; Mountain View, CA USA
ATCC	American Type Culture Collection; Rockville, MD USA
Amersham	Amersham; Arlington Heights, IL USA
Beckman	Beckman Instruments; Columbia, MD USA
Becton Dickinson	Becton Dickinson Labware; Lincoln Park, NJ USA
BIO 101	BIO101; La Jolla, CA USA
Biorad	Bio-Rad Laboratories; Hurcules, CA USA
Boehringer	Boehringer Mannheim Canada; Laval, PQ CAN
BRL	Life Technologies; Grand Island, NY USA
BTX	BTX; San Diego, CA USA
Chroma	Chroma Technology Corporation; Brattleboro, VT USA
Corel	Corel Corporation; Ottawa, ON CAN
Corning	Corning; Corning, NY USA
Fisher	Fisher Scientific; Pittsburg, PA USA
FMC	FMC BioProducts; Rockland, ME USA
Forma	Forma Scientific; Marietta, OH USA
Intelligenetics	IntelliGenetics/Betagen; Mountain View, CA USA
IEC	International Equipment Co.; Needham Heights, MA USA
MSI	Micron Separations; Westborough, MA USA
NEB	New England Biolabs; Mississauga, ON CAN
Nunc	Nunc Intermed; Roskilde, Denmark
Pharmacia	Pharmacia Biotech; Piscataway, NJ USA
Polaroid	Polaroid Corporation; Waltham, MA USA
Qiagen	Qiagen; Chatsworth, CA USA
Robbins	Robbins Scientific; Sunnyvale, CA USA
Sigma	Sigma Chemical, St. Louis, MO USA
Sorval	Sorval; Newtown, CT USA
Stratagene	Stratagene; La Jolla, CA USA
Tyler	Tyler Research Instruments; Edmonton, AB CAN
Vector	Vector Laboratories; Burlingame, CA USA
Zeiss	Carl Zeiss; West Germany

CHAPTER 1. INTRODUCTION

1.1. THE EUKARYOTIC GENOME

1.1.1. EARLY STUDIES

Much of our early understanding of the eukaryotic genome came from cytological studies. Feulgen staining was first used in the localization of DNA to the cell nucleus (Feulgen and Rossenbeck, 1924), and by 1934 Heitz had demonstrated that the genome consisted of two types of chromatin which were termed euchromatin and heterochromatin. Claude and Potter (1943) demonstrated that the chromatin threads observed in the nucleus were the chromosomes which served as a starting point for the *in vitro* isolation of eukaryotic chromosomes by Mirsky and Ris (1948).

Based on their observations of lampbrush chromosomes Callan and Lloyd (1960) and Gall and Callan (1962) were among the first to propose that the eukaryotic genome was composed of repeated gene sequences. Their hypotheses were based on the premise that each chromomere, or darkly staining band in the lampbrush chromosome, corresponded to the location of a gene. This thinking was supported by the fact that tritiated uracil and phenylalanine were incorporated into their respective macromolecules at looped outsites seen only at the chromomeres. Since each loop was larger than that expected to accommodate single genes, it was suggested that each gene was present in tandem copies.

The molecular analysis of eukaryotic genomes began in the late 1960s with the development of a standard method for determining the renaturation kinetics of total genomic DNA (Britten and Kohne, 1969). Construction of "Cot curves" generated by following the reassociation of eukaryotic DNA provides information on the reiteration and complexity of sequences in the genome. These analyses have permitted the identification of several major, distinct classes of genomic DNA that differed in their rates of reassociation: fast, intermediate, and slow. In most eukaryotes there is another fraction, which usually represents a few percent of the genome and anneals almost immediately. This very rapidly renaturing fraction, or snap back DNA, results from the presence of inverted repeats which can fold back upon themselves to form stem-loop and hairpin structures.

1.1.2. RAPIDLY REANNEALING SEQUENCES

Sequences of the rapidly reassociating class, which can comprise 5-15% of the genome depending on the species, are often tandemly repeated, highly reiterated, simple sequences (MacIntyre, 1994). These highly repetitive sequences range in length from 2-2200 base pairs with copy numbers ranging from 50 000 to about a million (Craig, 1994; MacIntyre, 1994). These sequences often differ in base composition from the bulk of the genome, thus allowing for their isolation by buoyant density centrifugation. The term satellite DNA was first used to describe a lower density band, mouse major satellite, found in addition to the bulk DNA resolved by CsCl centrifugation of mouse DNA (Flamm et.al., 1966). The classic chromosome *in situ* hybridization experiments of Pardue and Gall (1970) demonstrated that this mouse satellite DNA hybridized to DNA in the vicinity of the centromeric regions of the chromosomes. Although this major satellite has now been shown to lie outside the primary constriction, it has been implicated in at least some aspect of centromere structure and function (Lica et.al., 1986; Radic et.al., 1987). The satellite DNA families in the various mammalian orders appear to be unrelated in sequence, but posses similarities in their overall chromosomal organization, as well as short nucleotide stretches (Willard and Waye, 1987).

Telomeres also carry multiple copies of simple sequence repeats. Although there is variation in the sequence of the repeats from species to species, a consensus sequence of $C_{(1-8)}A/T_{(1-4)}$ is applicable to many species (Blackburn, 1984). A derivative of this consensus 5´ C_3TA_2 3´ has been found at all vertebrate telomeres (Meyne et al., 1989). The length of the telomeric repeats varies from species to species, and even from chromosome to chromosome within a species. Telomere lengths of as little as 20 bp up to several kilobases have been reported (Zakian, 1989).

1.1.3. INTERMEDIATELY REANNEALING SEQUENCES

Sequences in this class are moderately repetitive elements that are both tandemly repeated as well as interspersed among the unique sequences in the euchromatic chromosomal arms. This genomic component represents anywhere from 20-50% of the genome and contains members of large multigene families like the ribosomal RNA genes and the mini- and microsatellite repeats, but the major types of moderately repetitive elements are the short interspersed elements (SINEs) and long interspersed elements (LINEs).

Microsatellites are tandemly repeated sequences that are generally composed of 5 or fewer nucleotides. These simple sequence loci were first highlighted by Hamada and Kakunaga (1982), who demonstrated the

existence of hundreds of copies of the sequence $(TG)_n$ in yeast and tens of thousands in vertebrates. Tautz et al. (1986) demonstrated the existence of microsatellites comprised of a variety of simple sequences in several eukaryotes, and speculated that they arose through replication slippage. Microsatellites are widely distributed in mammalian genomes (Tautz and Renz,1984; Stallings et al., 1991; Beckmann and Weber, 1992) Microsatellite loci have proven to be invaluable in the construction of linkage maps, in a number of mammalian species, due in part to their high degree of polymorphism (Weissenbach et al., 1992; Deitrich et al., 1992; Jacob et al., 1995). Additionally, microsatellite sequences have been valuable in forensic cases (Hagelberg et al., 1991; Jeffreys et al., 1992), in parentage and relatedness tests (Amos et al., 1993; Morin et al., 1993) and in molecular population genetics (Bruford and Wayne, 1993). Expansion of microsatellite loci has been implicated in several clinical conditions, for example: expansion of $(CGG)_n$ in Fragile X syndrome (Fu et al., 1991) and expansion of $(CTG)_n$ in myotonic dystrophy (Brook et al., 1992).

Jeffreys et al. (1985) demonstrated the existence of longer repetitive elements termed minisatellites. The repeat units within minisatellites can be as large as 200 bp, and the allele size can be up to 50 kb (Bruford and Wayne, 1993). Minisatellite loci, unlike microsatellite sequences which apparently do not have a biased base composition, are predominately GC rich (Jeffreys et al., 1985; Nakamura et al., 1987). Minisatellites, at least in humans, have a restricted distribution and are predominantly located in the terminal regions of the chromosomes (Royle et al., 1988). Many minisatellite loci are extremely variable, with heterozygosity values greater than 90% and mutation rates of approximately 10^{-2} (Bruford and Wayne, 1993). This high degree of variation has made these loci an extremely valuable tool in molecular population genetic studies, genomic analysis and disease loci mapping. Microsatellites are preferred to minisatellites in many cases due to their ease of analysis by PCR, which allows for the detection of subtle differences in allele sizes.

SINES are short, moderately repetitive, interspersed sequences common among mammals, but also recognized in other organisms such as slime molds, locusts, fish and amphibians (Singer and Berg, 1991). SINES are generally between 100 and 500 bp in size and appear to be processed pseudogenes, originating from the amplification of RNA polymerase III transcribed sequences (Daniels and Deininger, 1985; Singer and Berg, 1991; Craig, 1994). SINES are often surrounded by target site duplications, and have a 3′ poly-A rich stretch. Each species has its own characteristic set of SINEs (Singer and Berg, 1991).

One of the best characterized SINE families is the Alu family of Old World primates and humans. Alu repeats occupy approximately 5% of the human

genome, with close to 10^6 dispersed copies (Schmid and Maraia, 1992). Alu repeats are found throughout the genome, in introns, in satellite sequences and in clusters with other interspersed repetitive sequences. Alu repeats are hetero-dimeric elements, bearing homology to regions of the internal promoter of 7SL RNA, which is involved in the translocation of newly synthesized proteins across the rough endoplasmic reticulum (Britten et al., 1988). Each monomer is about 130 bp and has a poly-A rich stretch at its 3′ terminus (Singer and Berg, 1991). The rodent Alu homolog, rodent type I family, is significantly different from the human Alu repeat, indicating that this SINE family must have evolved following the radiation of the mammalian orders (Singer and Berg, 1991). Additional rodent SINE families are homologous to various tRNA genes (Daniels and Deininger, 1985).

A model for the retrotransposition of SINEs has been proposed (Schmid and Maraia, 1992). The A and B boxes of a SINE element direct the initiation of RNA polymerase III transcription from a select, small number of SINEs. Transcription continues through the 3′ flanking direct repeat until it terminates in a stretch of 4 or more T residues. This results in a primary transcript in which the internal poly-A tract can base pair with the oligo U 3' terminus, generating an intramolecular primer for reverse transcription. Once reverse transcribed, the SINE DNA may insert at appropriate genomic loci. Only RNAs capable of taking on a tRNA-like structure are retrotransposed (Sinnett et al., 1991).

LINEs are long, interspersed, moderately repetitive elements found in mammals and a variety of other organisms including plants, fungi and invertebrates (Di Nocera and Sakaki, 1990). The most extensively studied LINEs in mammals are the primate Kpn I family and the Bam HI/MIF-1 family of rodents (Burton et al., 1986). These LINE families share homology, and are collectively known as the L1 family. There are approximately 10^4-10^5 L1 elements in mammalian genomes, representing as much as 10% or more of the genome (Burton et al., 1986; Pascale et al., 1993). While some L1 members are 6-7 kb in length, the vast majority of members are randomly truncated in their 5' ends resulting in an over-representation of the 3′ end in the genome (Voliva et al., 1983). Many elements possess a polyadenylation site and an A-rich stretch at the 3′ end and are flanked by direct repeats (Singer et al.,1983). These structural features suggest that L1 elements are insertions derived from a reverse transcribed L1 transcript (Rogers, 1985).

Full length L1 elements contain promoter-like sequences that are capable of directing transcription of heterologous sequences linked to them (Nur et al., 1988). The functionality of these promoter sequences is also supported by the fact that L1 elements are highly transcribed (Schmeckpeper et al., 1984). Furthermore, the elements contain two highly conserved open reading frames. The conserved nature of the two open reading frames

suggests they are required for the activity of the L1 element. Indeed, one of the open reading frames shares homology with reverse transcriptases, and the human protein product derived from the open reading frame has been demonstrated to function as a reverse transcriptase (Hattori et al., 1986; Mathias et al., 1991).

The chromosomal distribution of SINEs and LINEs appears to be complementary. Alu elements are concentrated in the light stained regions of G-banded chromosomes, while L1 elements are predominantly found in the dark stained regions of G-banded chromosomes (Korenberg and Rykowski, 1988). The light staining G-bands are also generally richer in genes than the dark G-bands (Korenberg and Rykowski, 1988).

1.1.4. SLOWLY REANNEALING SEQUENCES

This fraction comprises between 40 and 60% of the genome, and includes both unique genes and members of closely related gene families. Estimates of the total number of coding sequences in the mammalian genome indicate that there is likely some where in the range of 40 000-100 000 genes (Craig, 1994). The coding sequences are not uniformly distributed throughout the genome. The centromeric regions, comprised of simple sequence repeats, can be devoid of coding sequences over great lengths of tens of megabases (Southern, 1992). Conversely, the human class II and class III major histocompatibility locus has at least 80 genes within an 2 Mb region (Trowsdale et al., 1991). Spacing between exons shows a similar discontinuous distribution. There are clustered regions of the genome where the genes have small introns, while others have large intronic sequences. These regions seem to correspond to the light and dark bands observed on G-banded metaphase chromosomes (Bickmore and Sumner, 1989).

1.1.5. MAP CONSTRUCTION

1.1.5.1. GENERAL

The availability of a detailed map of all chromosomes of an organism, with the positions of genes accurately located, is a prerequisite to a complete understanding of genomic structure and gene function for an organism. The ideal map would contain genetic, physical and cytogenetic data, and be integrated such that information could be extracted at different levels of resolution. A variety of strategies and techniques presently utilized in the analysis of eukaryotic genomes are discussed in the following sections.

1.1.5.2. LINKAGE MAPS

Initial attempts to develop detailed genetic maps were based on Southern blotting analysis of restriction fragment length polymorphisms (RFLPs) revealed as the result of the gain or loss of restriction endonuclease recognition sites due to variation in simple sequences (Botstein et al., 1980). The first genetic map of the human genome, based on the segregation of 403 polymorphic loci in 21 reference families, had a resolution of 10-15 cM (Donis-Keller et al., 1987). For a number of reasons, including the labor intensiveness and the difficulty in distribution of probes for RFLP analysis, coupled with the advent of more polymorphic microsatellites has resulted in the development of newer methods for generating linkage maps.

Recently, emphasis has been placed on using the highly polymorphic and widely distributed microsatellite repeats as markers for developing linkage maps. Analysis of 814 $(CA)_n$ repeats segregating in 8 three generation human families has allowed for the generation of a linkage map spanning about 90% of the human genome with marker resolution of approximately 5 cM (Weissenbach et al., 1992). Theoretically, with an estimated 12 000 polymorphic $(CA)_n$ repeats in the human genome, it could be possible to generate a genetic map with a resolution of about 0.5 cM using only these markers (Weber, 1990). Similar mapping efforts have been carried out in the rat (discussed below; Jacob et al., 1995) and the mouse (Dietrich et al., 1992; Copeland et al., 1993). Detrich et al. (1992) constructed a linkage map of the mouse based on 317 simple sequence repeats having about 99% coverage and a marker density of 4.3 cM. The resolution of genetic maps is not likely to go much beyond the 1-2 cM level of resolution because of the effort required to find recombinants for this degree of resolution.

1.1.5.3. PHYSICAL MAPS

Physical maps include maps in the form of cloned contigs as well as maps in the form of long-range restriction and cytogenetic maps which position and order loci along a chromosome. Contigs are contiguous, overlapping fragments of cloned DNA that are arranged in an order that represents the sequences as they are found in the genome. Two strategies, top-down and bottom up, can be taken in mapping large genomes. The top down approach starts at the level of the intact genome and proceeds, for example by generating long range restriction maps using pulsed field gel electrophoresis of large DNA fragments, and assignment of DNA markers to the region. From the derived maps, cloned regions can be generated which can then be sequenced. The bottom up approach begins at the level of nucleotide sequence in a number of genomic clones and proceeds to assemble contigs by identifying overlapping regions in clones (Robinson, 1992).

The ability to clone large (greater than 30 kb) chromosomal fragments as single, contiguous pieces of DNA in a host vector has had a tremendous impact on the ability to generate physical maps of eukaryotic genomes. Vector systems that have proven most useful for the production of libraries of clones with inserts greater than 30 kb include cosmids, yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs) and the bacteriophage P1 vector system.

The most widely used approach to genome mapping involves the use of YAC cloning. YACs are plasmid based cloning vectors containing centromeric, telomeric, and autonomously replicating sequences which carry out the essential functions of yeast chromosomes (Burke et al., 1987). YAC cloning vectors can accommodate megabase size genomic DNA fragments (Adams et al., 1993). YAC libraries have been constructed from a wide variety of sources including total human genomic DNA (Brownstein et al., 1989; Traver et al., 1989; Anand et al., 1990), specific chromosomes or chromosomal regions (Little et al., 1989; McCormick et al., 1989; Wada et al., 1990; Abidi et al., 1990), *Caenorhabditis elegans* genomic DNA (Coulson et al., 1988), *Drosophila* genomic DNA (Garza et al., 1989), mouse genomic DNA (Larin et al., 1991), and *Arabdopsis* (Guzman and Ecker, 1988; Ward and Jen, 1990; Grill and Somerville, 1991).

Although YAC libraries provide the opportunity to recover clones covering large genomic regions, and would therefore require fewer clones to cover a genome, there are a number of disadvantages associated with them. In most libraries there is a high proportion, as high as 40-60%, of clones that are chimeric, containing non-contiguous genomic segments (Cooper, 1994). Additional problems associated with YACs include difficulty of transforming yeast, multiple YACs within the same yeast clone, instability of some cloned sequences and the presence of homologous widely distributed sequences which can complicate contig assembly (Bellanne-Cantelot et al., 1992; Schalkwyk et al., 1995). Analysis of YAC clones is also problematic. There are no easy methods for purifying YAC DNA from yeast genomic DNA, and colony hybridization methods are not as straight forward as bacterial screening methods due to the low copy number of YACs and the smaller number of cells in a yeast colony (Schalkwyk et al., 1995).

The recent development of intermediate capacity *Escherichia coli* based cloning systems, which have provided a more accessible and stable source of DNA than YACs, has contribute greatly to the construction of high resolution physical maps. One of these new cloning methods, the P1 cloning system, permits the cloning and propagation of genomic segments of 70-100 kb in a vector that contains features from the bacteriophage P1 (Sternberg, 1990). P1 vector cloning is similar to lambda-based cosmid cloning in that vector and genomic DNA molecules are ligated and packaged *in vitro* into a

phage particle. The phage is used to introduce the recombinant phage DNA into a host *Escherichia coli* strain expressing Cre recombinase, which acts on the loxP sites in the recombinant DNA molecule resulting in circularization. The circularized molecule is maintained at single copy in the cell by the P1 plasmid replicon, but the P1 lytic replicon can be induced with isopropylthio-β-D-galactoside to increase the copy number before preparing DNA. Another advantage of the P1 system is the positive selection system for clones containing inserts. Although the cloned inserts are smaller than those contained in YACs, the instability and chimerism observed with YACs has not been observed to an appreciable degree with the P1 cloning system (Sternberg, 1994).

P1 genomic libraries have been prepared from a variety of organisms and cell types including humans (Sternberg et al., 1990; Shepherd et al., 1994; Virgilio et al., 1993; Baxendale et al., 1993), *Drosophila* (Smoller et al., 1991; Lozovskaya et al., 1993), mouse (Pierce et al., 1992), rat (Southard-Smith and MacDonald, 1993), yeast (Hoheisel et al., 1993), pine tree (Gorman et al., 1992), crab (Shepherd and Smoller, 1994) and trypanosomes (Shepard and Smoller, 1994).

The utility of the P1 cloning system has already been demonstrated. P1 clones have been used in constructing contigs covering portions of the *hsp70/cps-1* locus in the mouse, the rat kallikrein locus (Gasser et al., 1993; Southard-Smith et al., 1994), and the human Huntington's disease gene (Baxendale et al., 1993). P1 clones also played a prominent role in the generation of a high resolution physical map of the 14 Mb genome of *Schizosaccharomyces pombe* (Hoheisel et al., 1993).

Another *Escherichia coli* based cloning system is the bacterial artificial chromosome, BAC (Shizuya et al., 1992). BACs use an F-factor based vector, which can accommodate genomic inserts up to 300 kb. Like the P1 system, replication of BACs is tightly controlled to maintain a low copy number, reducing the potential for recombination. BACs are introduced into *Escherichia coli* by electroporation which eliminates the size restriction which is imposed in systems which rely on packaging into a phage head. The BAC cloning system suffers by not having positive selection to identify clones with inserts.

More recently Ioannou et al. (1994) developed a P1 artificial chromosome (PAC) cloning system. The PAC cloning vector possesses features of both the P1 and F-factor based systems and can be introduced into *Escherichia coli* by electroporation. Insert sizes range from 100-300 Kb (Monaco and Larin, 1994).

1.1.5.4. CYTOGENETIC MAPS

The introduction of the present day FISH (fluorescence *in situ* hybridization) technology by Pinkel et al.(1986) has greatly advanced the field of genome analysis. This technique has permitted the rapid and routine assignment of a large number of cosmid, YAC and other clones to specific chromosomal regions (for example: Lichter et al., 1990; Wada et al., 1990). The development of multiple labeling and detection systems has allowed for the simultaneous mapping and ordering of multiple sequences with high resolution (Lichter et al., 1990; Trask et al., 1991; Reid et al., 1992; Dauwerse et al., 1992). The power of FISH in constructing maps is demonstrated in a study by Lengauer et al. (1993) in which by carefully choosing fluorochromes with complementary excitation and emission spectra, and using more than one type of labeled nucleotide in some probes these researchers were able to simultaneously localize up to 20 different YACs on the same human chromosome.

The resolution of mapping genes, or DNA segments with conventional FISH methods to metaphase chromosomes is in the range of 1-2 Mb (Buckle and Kearney, 1994). Improvement of the resolution observed in standard FISH to metaphase chromosomes has been achieved by applying FISH to interphase cell nuclei where the DNA is less condensed, allowing for the distinction of probes separated by 100-500 kb (Trask et al., 1989; Lawrence et al., 1990; Van den Engh et al., 1992. Several new high resolution FISH mapping protocols that utilize decondensed chromatin, and naked DNA fiber targets have recently been described (Heng et al., 1992; Weigant et al., 1992; Parra and Windle, 1993; Houseal et al., 1994; Fidlerova et al., 1994; Heiskanen et al., 1994). All of these high resolution Fiber-FISH techniques are based on the release of chromatin fibers from the nuclear scaffold and their subsequent use as hybridization targets.

Weigant et al. (1992) released DNA by detergent treatment such that highly extended DNA loops were arranged around the nuclear matrix in a halo-like structure. The resolution achieved with this method is 10-200 kb. Heng et al. (1992) have used various drug treatments and a high pH buffer to release extended chromatin fibers from cells such that sequences as close as 20 kb could be distinguished from one another. Sequences spread over 350 kb could be ordered. The DIRect VISualization Hybridization (DIRVISH) protocol described by Parra and Windle (1993) provides the added benefit that the released DNA is physically stretched by tilting the slide such that linear stream of DNA fibers are produced to serve as targets in hybridization. Overlapping probes as small as 5 kb can be detected with the DIRVISH protocol (Parra and Windle, 1993). Heiskanen et al. (1994) have described a method for generating extended DNA hybridization targets from melted pulsed field gel electrophoresis blocks and physical stretching with a glass

slide run over the released DNA. Multiple probes as small as 4 kb could be detected over a linear range of 500 kb.

These new cytogenetic mapping methods provide a means to visually inspect the arrangement of specific sequences along a stretched single DNA fiber. Application of these techniques have already begun to assist in the orientation of clones, assess the degree of clone overlap, estimate the size of uncloned gaps between adjacent contigs and in the examination of intragenic organization and rearrangements (Bengtsson et al., 1994; Tocharoentanaphol et al., 1994; Heiskanen et al., 1995; Pizzuti et al., 1996).

1.1.6. THE RAT GENETIC MAP

There are over 140 inbred strains of the laboratory rat, *Rattus norvegicus*. Many have been developed for specific disease characteristics such as cancer, diabetes, hypertension, cavity formation, alcoholism and various immunological defects. The rat has clearly established itself as the most widely studied experimental animal system for biomedical research (Jacob et al., 1995). Despite its wide-spread use for biomedical research, which dates back to before 1850 (Lindsey, 1979), the rat has lagged behind the mouse as a genetic model organism. Although the rats larger size has been viewed as an advantage for biomedical research, it has been a disadvantage for furthering genetic analysis of this species due to the greater space required for housing the large number of animals required for such analyses. The ever increasing convergence of the fields of molecular biology with the biomedical sciences has resulted in an increased need for genetic studies in the rat. A prerequisite to comprehensive genome analysis of any organism is the availability of a detailed genetic map.

The first complete genetic linkage map of the laboratory rat was made available in 1995 (Jacob et al., 1995). The map is based on 432 simple sequence length polymorphisms genotyped from a single F_2 intercross and anchoring of identified linkage groups to specific rat chromosomes using some 550 loci which had previously been assigned chromosomal locations using somatic cell hybrid panels.

The new map provides coverage of the entire genome with linkage groups on all 21 rat chromosomes with an average marker spacing of 3.7 cM. This map serves as a starting point from which further rat genome analyses such as genome scans for genes underlying medically relevant traits and complete physical mapping can proceed.

1.2. MULTIGENE FAMILIES

1.2.1. INTRODUCTION

Multigene families are defined as groups of genes with sequence homology and possessing overlapping functions (Hood et al., 1975). Multigene families are widespread in the genomes of higher eukaryotes. All multigene families are postulated to have arisen by an initial duplication event. Gene duplication is viewed as one of the primary mechanisms for the evolution of new gene functions, and increased genomic and organismic complexity (Walsh, 1995; Graham, 1995).

Four types of events have been suggested to lead to the gene duplication required for the formation of a multigene family: non-homologous chromosomal breakage and reunion, unequal, but homologous recombination between two copies of a repeated sequence that fortuitously lies on either side of the region to be duplicated, RNA-mediated transposition and gene amplification resulting from over-replication (Maeda and Smithies, 1986). Analysis of duplicated sequences in the eukaryotic genome has provided evidence for each of these mechanisms (Maeda and Smithies, 1986).

A common characteristic of multigene families is the lack of the expected degree of divergence between members. Assuming base changes to be neutral, family heterogeneity within a species is expected to be the same as heterogeneity between two species (Dover et al., 1982). However, analysis of many multigene families has clearly indicated a higher degree of family homogeneity within a species than there is for the same family between species (Dover et al., 1982). The pattern of within-species homogeneity and between-species heterogeneity for a family of repeated sequences is known as concerted evolution (Dover, 1982). When functional diversity is required by an organism, concerted evolution results in gene families with variable members, whereas more homogeneous gene families arise when large amounts of a uniform gene product are required (Ohta, 1988). Examples of both of these outcomes are evident in the eukaryotic genome.

1.2.2. GENOMIC ORGANIZATION OF MULTIGENE FAMILIES

Some, and probably the majority, of the mechanisms for the duplication of genes leave the duplicated genes on the same chromosome, often within a few kilobases of each other. Gene families fall into three commonly recognized organizational patterns: dispersed, clustered and tandemly arranged genes.

Dispersed gene families (Table 1.1) are composed of genes that are scattered among non-homologous chromosomes, or are far enough apart on the same chromosome such that their syntenic relationship does not increase interactions between members. Examples of dispersed functionally related gene sets are also evident in the eukaryotic genome. For example, in *Saccharomyces cerevisiae*, the core histone protein genes are arranged as dispersed, divergently transcribed gene pairs, with H2A paired with H2B and H3 paired with H4 (Smith, 1984). The core histone genes in *Caenorhabditis elegans* are similarly organized, with dispersed sets containing at least one copy of each of the four core histones (Roberts et al., 1987).

Members of clustered gene families (Table 1.1) are linked with neighboring genes, often within a few kilobases of each other, but display an irregular spacing interval between genes (Graham, 1995). The genes are often inverted with respect to each other in an irregular pattern. In addition to the conserved genic regions, there are regions surrounding the genes that are also conserved between members of the family. The genes are linked by non-conserved DNA sequences.

Tandemly arrayed genes (Table 1.1) are located within DNA segments that are repeated in a head-to-tail manner a number of times. The repeated DNA segment can contain more than a single gene. In cases where multiple genes are present within the repeated segment, the orientation and order of genes is constant for each repeated unit (Graham, 1995). Arrays with oppositely oriented repeating units are rare, probably because unequal crossing over within such an array would result in duplications and deletions which would prove detrimental (Graham, 1995).

Tandemly arranged and clustered gene families differ in a number of respects. Not only does their physical organization differ, but they differ in the interactions that occur between members, in the benefits they provide to the host genome, in the likelihood of generating new genetic complexity and in their durability under conditions of relaxed selection (Graham, 1995).

1.2.3. EVOLUTION OF MULTIGENE FAMILIES

Dispersed genes, as a consequence of the physical distance separating family members, are not subject to the influences of unequal crossing over. Similarily, the effects of gene conversion are on dispersed gene families is diluted due to the distance between family members (Graham, 1995). One result of the reduced influence of these two evolutionary mechanisms is that sequences surrounding individual members of dispersed gene families are unrelated (Morzycka-Wroblewska et al., 1985). The family members themselves are also more heterogenous than those of the other two main organizational families (Graham, 1995).

Table 1.1. Examples of Multigene Families and Their Organization

Gene Family	Species	Gene Number	Organization	Reference(s)
Histones				
	Homo sapiens	20-40	clustered	1
	Mus musculus	20	clustered	2, 3
	Gallus domesticus	10-20	clustered	4
	Xenopus laevis	40-50	tandem and clustered	3, 5
	Diptera	100-200	tandem	6-8
	Caenorhabditis elegans	11	dispersed sets	9
Actin				
	Homo sapiens	20-30	dispersed	10
	Rattus	≥8	dispersed	10
	Gallus domesticus	8-10	dispersed	10
	Drosophila	6	dispersed	10
	Dictyostelium	17	dispersed and clustered	10
U1 snRNA				
	Homo sapiens	30	tandem	11
	Rattus	50	clustered	12
	Mus musculus	10	clustered	13
	Xenopus laevis	500	tandem and clustered	13, 14
U2 snRNA				
	Homo sapiens	10-20	tandem	15, 16
	Mus musculus	10	clustered	17
	Gallus domesticus	35-40	tandem	18
	Xenopus laevis	500-1000	tandem	13, 19

Table 1.1. continued Examples of Multigene Families and Their Organization

Gene Family	Species	Gene Number	Organization	Reference(s)
5S rRNA				
	Homo sapiens	2000	tandem	10
	Drosophila melanogaster	160	tandem	10
	Neurospora crassa	100	dispersed	20, 21
	Xenopus laevis			
	oocyte	20 000	tandem	10
	somatic	400	tandem	10
18-5.8-28S rRNA				
	Homo sapiens	200	dispersed tandem repeats	22
	Rattus norvegicus	160	dispersed tandem repeats	22
	Mus musculus	100	dispersed tandem repeats	22
	Saccharomyces cerevisiae	140	tandem	22
Type-I Interferon				
	Homo sapiens	26	tandem	23
γ-Crystallins				
	Rattus	6	tandem	24
Type-I Keratins				
	Mus musculus	13	likely tandem, but definitely clustered	25
	Homo sapiens	13	likely tandem, but definitely clustered	26
Kallikreins				
	Mus musculus	24		27
	Homo sapiens	3		28
	Rattus	15-20		29

The table provides information regarding the organization of several multigene families in the indicated species. References: (1) Stein et al., 1984; Marzluff and Graves, 1984; (3) Maxson et al., 1983; (4) D' Andrea et al., 1985; (5) Van Dongen et al., 1984; (6) Colby and Williams, 1993; (7) Kremer and Hennig, 1990; (8) Hankeln and Schmidt, 1991; (9) Roberts et al., 1987; (10) Singer and Berg, 1991; (11) Bernstein et al., 1985; (12) Watanabe and Ohshima, 1988; (13) Dahlberg and Lund, 1988; (14) Zeller et al., 1984; (15) Lindgren et al., 1985; (16) Van Arsdell and Weiner, 1984; (17) Nojima and Kornberg, 1983; (18) Korf and Stumph, 1986; (19) Mattaj and Zeller, 1983; (20) Morzycka-Wroblewska et al., 1985; (21) Selker et al., 1981; (22) Long and Dawid, 1980; (23) Diaz et al., 1994; (24) den Dunnen et al., 1986; (25) Filion et al., 1994; (26) Barletta et al., 1990; (27) Evans et al., 1987; (28) Evans et al., 1988; (29) Southard-Smith et al., 1994

Members of dispersed gene families, because of their isolation, are more likely than members of clustered and tandemly arrayed gene families to become distinct genes and evolve new functions (Graham, 1995).

Clustered gene families are usually not subject to unequal crossing over, but are influenced by gene conversion (Graham, 1995). The combined action of natural selection, mutation and conversion result in the observed islands of homology surrounding clustered genes, while at the same time allow for the presence of highly divergent sequences both upstream and downstream of the islands. Natural selection acts to eliminate many of the changes that may occur within the gene and nearby adjacent regions, while mutations are permitted to accumulate in the linker regions. Conversion events are restricted to regions displaying similar sequences, and are greatest in regions displaying the greatest degree of homology (Waldman and Liskay, 1987; McCormack and Thompson, 1990). As a result conversion events will continually occur in regions where they have previously taken place ensuring regions of high similarity. The non-random distribution of conversion events, coupled with the fact that conversion events can be very short, often as few as 10 bp in length contribute to the sharp contrast between the conserved islands of homology and the divergent upstream and downstream sequences (Wheeler et al., 1990). Transcribed sequences also appear to be converted preferentially over non-transcribed regions (Voelkel-Meiman and Roeder, 1990).

Clustered gene families are usually more diverse than tandemly arrayed gene families. One reason for this outcome is that conversion can selectively homogenize portions of genes in clustered families, leaving the rest of the gene to diverge. This is typified by the lysozyme genes of the mouse. The two differentially expressed types were generated by duplication followed by conversion between parts of the two genes (Cross and Renkawitz, 1990). A similar mechanism is proposed to be responsible for the evolution of the homeotic genes in *Drosophila melanogaster* (Deatrick et al., 1991). A second reason for the increased diversity of clustered gene families is that numerous small conversion events acting in select regions of a large cluster of genes can generate new sequence combinations (Graham, 1995). This type of shuffling of sequences has been observed at the chorion locus in *Bombyx mori* (Xiong et al., 1988), and mammalian major histocompatibility loci (Baltimore, 1981).

Tandemly arrayed genes are subject to the homogenizing effects of unequal crossing over. Unequal crossing over acts to homogenize the entire repetitive element, and not just gene sequences (Arnheim, 1983). Conversion events also affect tandem arrays (Roberts et al., 1984). Although genes within tandem arrays may evolve, they can not do so independently. Therefore, unlike clustered genes, tandemly arrayed genes are limited in

their capacity to diverge (Ohno, 1970).

1.2.4. RELATIONSHIP BETWEEN GENE NUMBER AND ORGANIZATION

There is a strong relationship between the size of a gene family and its organization. Gene families with more than approximately 50 members are almost always organized as tandem repeats, while families with 10-50 members are usually clustered, but occasionally show a tandem arrangement and smaller gene families display a dispersed arrangement (Graham, 1995).

It would appear that only organization in tandem arrays will provide for homogenization rates high enough to maintain large multigene families. In the absence of homogenizing mechanisms, a multigene family would degenerate until it contained the minimum number of genes required for survival. Allowing a gene array to degenerate to the minimum number of genes makes the species more sensitive to mutation, changes in environmental conditions, and may impose a high cost to outbreeding (Tschudi et al., 1982; Graham, 1995). Tandem gene arrangements may also serve to maintain a balance between genes that are required to be present in a particular stoichiometric ratio (Thomashow et al., 1983; Meeks-Wagner and Hartwell, 1986; Clark-Adams et al., 1988; Drouin et al., 1992).

The clustered gene arrangement characteristic of smaller gene families serves to protect them against fluctuations in gene numbers caused by unequal crossing over. The loss of even a single gene from a small gene family could cause difficulty for the organism (Graham, 1995). Reduction of a gene family to a single repeat, which could be possible if a small gene family was subject to unequal crossing over, would prevent its re-expansion by means other than meiotic recombination.

1.2.5. RELATIONSHIP BETWEEN ORGANIZATIONAL ARRANGEMENTS

The observation that the same gene family can possess a tandem arrangement in some species and a clustered organization in a closely related species, indicates that tandem arrays and clusters share a recent ancestry. For example the U1 snRNA genes are tandemly arranged in humans, but are clustered in mice (Bernstein et al., 1985; Dahlberg and Lund, 1988). Several examples of gene families with different organizations coexisting in the same genome also support a close relationship between clustered genes and tandemly repeated genes (Maxson et al., 1983). These observations suggest that either tandem arrays can give rise to clusters, that clusters can give rise to tandem arrays, or that dispersed genes can give rise

to both of the other organizational patterns (Graham, 1995).

Several lines of evidence indicate that a clustered gene family could arise from a tandemly arrayed gene family. One means by which this transformation could take place is by generating a dispersed gene as an intermediate which then evolves into a cluster. Tandem and dispersed organization of the same gene family in closely related species indicates that transitions between these two patterns of organization can occur fairly rapidly (McCutchan, 1986; Guay et al., 1992; Matera et al., 1990). It appears that the transition is bi-directional. Tandem arrays can evolve from a dispersed gene by a variety of duplication mechanisms (Maeda and Smithies, 1986), and dispersed "orphan" genes derived from tandem arrays have been described (Childs et al., 1981). The second step in the transformation of a tandem array to a cluster would be the evolution of the cluster from the newly formed dispersed gene. Another mechanism proposed for the generation of a clustered gene family from a tandemly arrayed set of genes is through a process of duplication, divergence and further duplication of a head-to-tail tandem repeat. Such a model has been suggested to be responsible for the evolution of the human alpha-globin locus and the ovine beta-globin cluster (Liebhaber et al., 1981; Garner and Lingrel, 1988; De Angioletti et al., 1992).

A major distinction between tandemly arrayed gene families and clustered gene families is that the former under goes unequal crossing over. The transformation of a tandem array to a clustered arrangement would require the cessation of unequal crossing over, and the maintenance of conversion. One means by which this selective suppression is achieved is by inversion of one or more of a tandem array (Graham, 1995). Unequal crossing over is not compatible with an inverted gene organization as it would lead to duplications and deletions which may have an adverse effects on the organism (Graham, 1995).

It also appears that amplification of small parts of gene clusters to form new tandem arrays is also possible. Evidence for this comes in the form of tandem repeats bearing regions characteristic of clustered genes (Berger et al., 1978; Lund et al., 1984; Scott et al., 1988; Xiong et al., 1988).

1.3. THE LIPOCALINS

1.3.1. STRUCTURE OF LIPOCALIN PROTEINS

Unlike multigene families, which are defined as groups of genes with sequence homology and related overlapping functions (Hood et al., 1975), supergene families are broader groups of genes that share one or more domains of common origin (Dayhoff, 1978; Doolittle, 1981). The lipocalins

are a recently identified supergene family first proposed with the resolution of the three-dimensional structure of retinol-binding protein (Newcomer et al., 1984).

It has become apparent that the lipocalins are encoded by a large, diverse supergene family. Lipocalins are typically small (160-190 amino acid residues) proteins that possess a high affinity for binding hydrophobic molecules (see Table 1.2) (Pervaiz and Brew, 1985; Flower et.al., 1993). Until recently, lipocalins had only been found in eukaryotic organisms, mostly vertebrates, although some have been identified in arthropods (Keen et al., 1991a, b; Granforninia et al., 1995). More recently, lipocalins have been identified from several bacterial species suggesting an ancient origin for this supergene family (Flower et al., 1995).

The crystallographic structures of several lipocalins have now been determined (North, 1991). Comparisons indicate several structurally conserved regions within the proteins. The lipocalin structure (see Figure 1.1) is characterized by eight β-strands which form a cup-shaped antiparallel β-barrel largely responsible for the sequestration of the ligand. The interior of the β-barrel is lined with a high proportion of hydrophobic residues, making the lipocalins ideally suited for binding and transporting hydrophobic ligands. The eight strands of the barrel are linked by a succession of β-hairpin loops, with the exception of linking loop 1 which forms a large Ω loop responsible for capping the ligand binding site. A short 3_{10} - helix at the N-terminus directly before the first β-strand is responsible for closing off the other end of the barrel. C-terminal to the last β-strand is an α-helical region which packs up against the last two β-strands. The extreme C-terminus is characterized by a short β-strand which participates in β-sheet hydrogen bonding with the first β-strand of the barrel.

Flower et al. (1993) analyzed four lipocalins and have demonstrated that the strands of the β-barrel are well conserved and that most of the structural variation occurs in the intervening loops, which differ in both size and conformation. The variation in the loops suggests that they are involved in the functional properties of individual member proteins (Sivaprasadarao et al., 1993; Flower, 1995). Furthermore, within the barrel, there is variation with respect to the degree of conservation. The greatest structural preservation is observed for the first, sixth, seventh and eighth β-strands of the barrel, while the remaining strands of the barrel form shorter structurally conserved regions. The C-terminal β-strand is also part of the structurally conserved core, as is the N-terminal 3_{10}-helix. Although the α-helical region C-terminal to the barrel is conserved in all lipocalins examined, it does not form part of the structurally conserved core since neither its length or position relative to the core is maintained between proteins.

Table 1.2. Representatives of the Lipocalin Superfamily of Proteins

Protein	Size	Tissue(s)	Fluid	Ligand(s)	Function	References
β-lactoglobulin	162	mammary gland of ruminants	milk	retinol/hydrophobic molecules	vitamin transport to gut	1, 2
α-1-microglobulin	183	liver (mammals)	serum, urine cerebrospinal fluid	IgA/retinoids/porphyrin	mediation of neutrophil proliferation and chemotaxis	3-5
α-1-acidglycoprotein	187	liver (mammals)	serum, urine	progesterone	mediates inflamatory response	6-10
purpurin	175	retina (chicken)		retinol/heparin/glycosamino-glycan	cell adhesion and survival	11-13
crustacyanin	174	carapace (lobster)		astaxanthin	coloration/photoreception	14, 15
apolipoprotein D	169	adrenal, kidney, pancreas, liver, intestine, apocrine (mammals)	serum, gut secretions apocrine secretions	lecithin/cholesterol/E-3-methyl-2-hexanoic acid	lipid transport/tissue repair/ chemical signalling	16, 17
insecticyanin	189	larval fat body (hornworm, butterfly)	haemolymph	biliverdins	camouflage/photoreception	18, 19
alpha-2u-globulin/MUPs	162	liver, salivary, mammary, preputial lachrymal glands (rat/mouse)	serum, urine, tears, saliva	pheromones	chemical signalling	20-22
retinol binding protein	183	liver (mammals)	serum	retinol/transthyretin	retinol transport	23, 24
human placental protein 14	162	placenta, secretory endometrium	amniotic fluid	?	immunopressive activity during implantation and placentation	25, 26

Table 1.2. continued Representatives of the Lipocalin Superfamily of Proteins

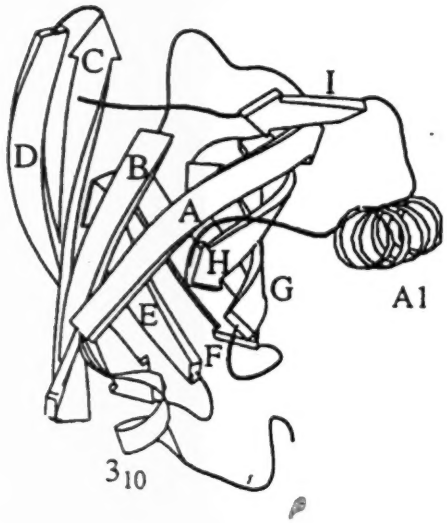
Protein	Size	Tissue(s)	Fluid	Ligand(s)	Function	References
Tear lipocalin/Von Ebner's gland protein	176	lachrymal/Von Ebner's glands (mammals)	tears/saliva	retinol/cholesterol/fatty acids and alchols/glyco- and phospho-lipids	tear film formation/taste perception	27-29
androgen-dependent epididymal protein	165	epididymis (rat)	epididymal fluid	?	sperm maturation	30
neutrophil gelatinase associated protein	179	neutrophils, epithelial cells		gelatinase/bacterial formyl peptides	modulation of gelatinase activity/ anti-inflammatory function/ immuno-modulatory function	31-33
aphrodisin	152	vagina	vaginal discharge	pheromones	chemical signalling	34-36
prostaglandin D synthetase	190	central nervous system, retina, cochlea		prostaglandin H_2/ bilirubin/ indocyanine green/ 1-chloro-2,4-dinitrobenzene	maturation and maintenance of the central nervous system	37, 38
complement component-8 γ subunit	182			C8 binding protein on erythrocyte cell surface/ retinol	protection against cell lysis induced by the membrane attack complex of complement	39, 40

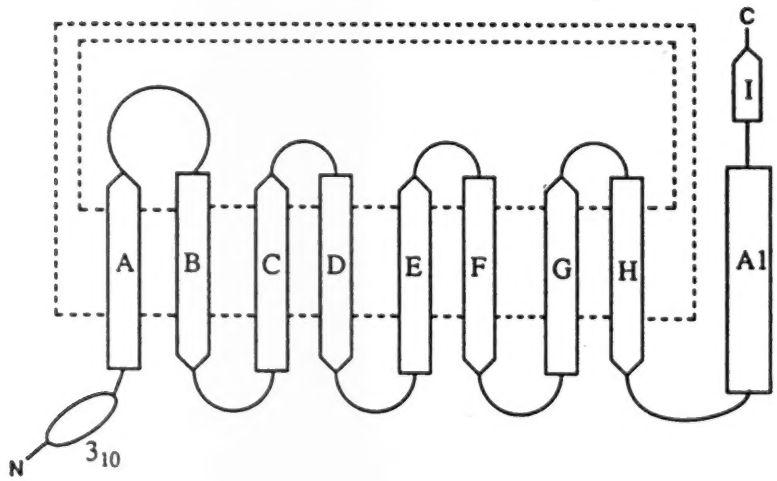
The table indicates the number of amino acid residues making up the mature polypeptide of several lipocalins in addition to, when known, the tissues the proteins are expressed in, the fluids the proteins are secreted in, known ligands for the proteins, and proposed functions for the specified lipocalin. References:
(1) Papiz et al., 1986; (2) Monaco et al., 1987; (3) Logdberg and Akerstrom, 1981; (4) Mendez et al., 1986; (5) Tehler et al., 1978; (6) Liao et al., 1985; (7) Baumann and Held, 1981; (8) Baumann et al., 1983; (9) Dente et al., 1987; (10) Ganguly et al., 1967; (11) Flower, 1995; (12) Shubert and LaCorbiere, 1985; (13) Shubert et al., 1986; (14) Keen et al., 1991; (15) North, 1995; (16) Drayna et al., 1986; (17) Zeng et al., 1996; (18) Riley et al., 1984; (19) North, 1991; (20) Finlayson et al., 1965; (21) Roy, and Neuhaus, 1966a; (22) Bocskei et al., 1992; (23) Gudas, 1994; (24) Bloomhoff et al., 1990; (25) Van Cong et al., 1991; (26) Julkunen et al., 1988; (27) Redl et al., 1992; (28) Kock et al., 1994; (29) Glasgow et al., 1995; (30) Brooks et al., 1986; (31) Kjeldsen et al., 1994; (32) Bundgaard et al., 1994; (33) Kjeldsen et al., 1993; (34) Singer et al., 1986; (35) Henzel et al., 1988; (36) Singer and Macrides, 1990; (37) Nagata et al., 1991; (38) White et al., 1992; (39) Hansch, 1988; (40) Sodetz, 1988.

Figure 1.1. The lipocalin structure.

(A) Ribbon diagram of mouse urinary protein. The β-strands are labeled A-I. The 3_{10}-helix before strand A and the α-helix (A1) beyond strand H are also labeled.

(B) Schematic representation of the lipocalin fold. The nine β-strands are indicated by arrows and labeled A-I. The N-terminal 3_{10}-helix and the C-terminal α-helix (A1) are also labeled. Hydrogen bonding between β-strands is indicated by dashed lines. The connecting loops are shown as solid lines. From Flower (1994).


C
I
D
B
A
H
G
A1
E
F
3_{10}


C
I
A
B
C
D
E
F
G
H
A1
N
3_{10}

1.3.2. LIPOCALIN PROTEIN SEQUENCE CONSERVATION

Lipocalins generally exhibit a low level of amino acid residue identity, often in the range of 10-30% in pairwise comparisons (Flower et.al., 1993). The most conserved amino acid motifs are a Gly-Xaa-Trp-Y(Y represents an aromatic or basic residue) sequence near the amino terminus, a Thr-Asp-Tyr-Xaa-Xaa-O (O represents an amino acid with an aromatic side chain-almost always a tyrosine) in the central portion of the protein, and two cystein residues at conserved positions.

When examining the sequences of lipocalins which have been designated inter-species homologues based on function and expression patterns, a wide divergence is apparent. For example the mouse major urinary proteins and their rat homologue, the alpha-2u-globulins are 66% identical (Clark et.al., 1984a); Human and rat α-1-acid glycoproteins are 44% identical; and human and rat Von Ebners gland protein are 60% identical (Kock et.al., 1994). In striking contrast to the interspecific variation, is the remarkable intraspecific conservation of the sequences of duplicated genes in lipocalin subfamilies. Six mouse major urinary protein genes are 95% identical (Clark et.al., 1984a); mouse α-1-acid glycoproteins 1 and 2 are 81% identical; and human Von Ebner glands proteins 1 and 2 share 98% identity at the DNA sequence level (Kock et.al., 1994).

Igarashi et.al. (1992) generated a phylogenetic tree using the amino acid sequences of 25 members of the lipocalin superfamily (see Figure 1.2). The tree displays clustering consistent with gene organization and cysteine residue positions. Many of these genes have a characteristic gene structure of seven exons with a 5′ untranslated region in the first exon and an entirely non-coding seventh exon (Igarashi et.al., 1992). The retinol binding protein and apolipoprotein D genes, both of which exhibit fusion of their downstream exons when compared to the other lipocalins, form a cluster apart from the other lipocalin genes. Within this cluster apolipoprotein D, the bilin binding proteins, and the crustacyanins form a sub-cluster of genes which have a unique cystein residue positioning (aligned amino acid positions 73 and 220, compared to positions 109 and 220 for the other lipocalins).

Sansom et al. (1994) applied a three-dimensional structural searching method to generate structural sub-families within the lipocalins. In most instances the structural subclasses corresponded exceedingly well with the gene clusters observed by Igarashi et.al. (1992) using sequence similarities.

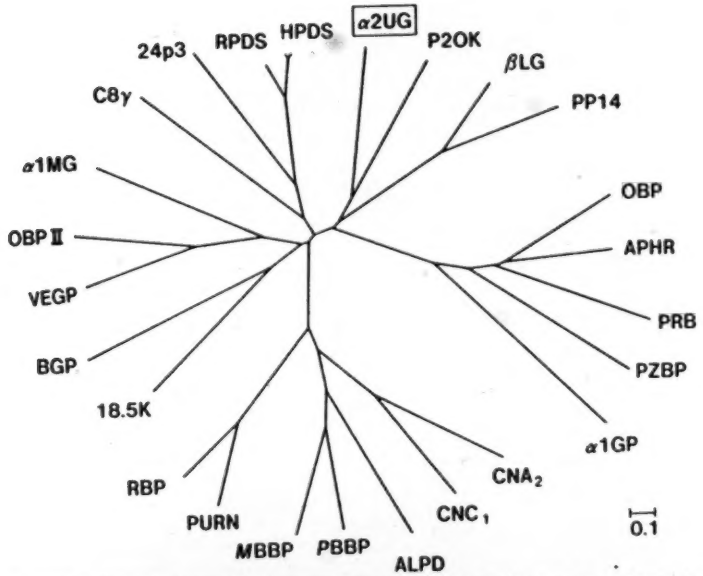


Figure 1.2. Phylogenetic tree of 25 members of the lipocalin family. Scale bar represents branch length corresponding to 0.1 amino acid substitutions per site. From Igarashi et al. (1992). Abbreviations:
RPDS, rat brain prostaglandin D synthase
HPDS, human brain prostaglandin D synthase
24p3, mouse 24p3 protein
C8γ, human complement component 8 γ chain
α2UG, rat alpha-2u-globulin
p20K, chicken quiescence-specific polypeptide 20K
βLG, bovine β-lactoglobulin
PP14, human placental protein 14
OBP, rat odorant-binding protein
APHR, hamster aphrodisin
PRB, rat probasin
PZBP, bovine pyrazine-binding protein
α1GP, human α-1-acid glycoprotein
CNA_2, lobster crustacyanin A_2
CNC_1, lobster crustacyanin C_1
ALPD, human apolipoprotein D
PBBP, butterfly bilin-binding protein
MBBP, tobacco hornworm bilin-binding protein
PURN, chicken purpurin
RBP, rat retinol-binding protein
18.5K, rat androgen-dependent epididymal 18.5K protein
BGP, frog Bowman's gland protein
VEGP, rat von Ebner's gland protein
OBP II, rat odorant-binding protein II
α1MG, human α-1-microglobulin

A few anomalies were observed: the γ chain of human complement protein C8 is nearer in sequence to rat alpha-2u-globulin, but seems structurally related to the bilin binding proteins, and α-1-acid glycoproteins and crustacyanin C1 would not be predicted to be lipocalins based on the structural search method alone.

Although the overall amino acid sequence similarity between family members is low, small regions of amino acids that are nearly invariant between family members, is sufficient to allow for the classification of a protein as a lipocalin. These small regions of amino acid identity are significant enough to ensure a common three-dimensional folding pattern designed for binding and transporting hydrophobic molecules.

1.3.3. LIPOCALIN FUNCTION

For many of the lipocalins, the biologically relevant ligands are not known. Additionally, it is not clear, given the structural similarities of the lipocalins, whether they bind specific ligands, or bind a diverse set of hydrophobic molecules. Presently, evidence suggests that both cases exist. For example, retinol-binding protein is quite specific for retinol (Kanai et al., 1968), whereas the odorant binding proteins are more indiscriminating in their binding specificities (Pelosi, 1994). The genes encoding lipocalins are expressed in a variety of secretory tissues, and the resulting proteins are secreted into a wide variety of bodily fluids such as plasma, urine, cerebrospinal fluid, nasal mucus, amniotic fluid, milk, saliva and tears. Other lipocalins are found inside cells with either a cytoplasmic or nuclear localization.

Although lipocalins have generally been considered to be transport proteins, it is becoming increasingly apparent, as demonstrated by the proposed functions for the family members presented in Table 1.2, that this supergene family participates in a much wider range of biological functions, including not only the binding and transport of hydrophobic molecules, but also the formation of macromolecular complexes and the binding to specific cell-surface receptors to initiate signal transduction pathways. Although the physiological roles of all the lipocalins have not been elucidated, especially those identified on the basis of sequence homology alone, a common theme in the function of many lipocalins appears to be a role in cellular regulation and homeostasis (Flower 1994). For example, alpha-1-microglobulin and neutrophil gelatinase associated lipocalin are mitogenic, acting to enhance cell growth; whereas apolipoprotein D and purpurin act to stabilize or maintain mature cell populations. Yet other lipocalins like odorant binding proteins, aphrodisin, alpha-2u-globulin, MUP, and Von Ebner's gland protein act to sequester and safely transport biologically significant molecules in a controlled manner. It is clear, in many cases, that lipocalins mediate this

regulatory role through various interactions with cellular receptors.

It has been suggested that designer proteins based on the structural properties of the lipocalins may be of some medical and pharmaceutical import (Godovac-Zimmermann, 1988). Given the interactions of several lipocalins with specific cell surface receptors, it may be feasible to manipulate ligand binding specificity without affecting receptor affinity, so as to target drugs to certain tissues or to alter the levels of endogenous hydrophobic metabolites such as cholesterol. Since free fatty acids have been shown to possess antimicrobial properties (Miller et al., 1988) the role of lipocalins as carriers of endogenous antimicrobial molecules presents itself as another possibility (Glasgow et al., 1995).

1.3.4. LIPOCALIN GENOMIC ORGANIZATION

Comparison of the exon/intron organization of the lipocalins demonstrates that the majority of lipocalin genes contain six protein-coding exons and a non-coding seventh exon. Additionally, the first exon of many lipocalins contains coding and non-coding sequences. The intron size and phasing is also highly conserved among the various members of the family, suggesting the lipocalins arose from the duplication of a common ancestral gene (Holzenfeind and Redl 1994).

Many of the lipocalin proteins are encoded by multiple genes. Examples of this phenomenon include the cluster of three human alpha-1-acid glycoprotein genes, the array of 35-40 MUP genes and 20 alpha-2u-globulin genes in the mouse and rat respectively, and the multiple odorant binding proteins observed in many species among others.

Seven genes coding for lipocalins in humans (alpha-1-microglobulin, tear lipocalin/Von Ebner's gland protein, prostaglandin D2 synthase, neutrophil gelatinase associated lipocalin, placental protein 14, complement component 8γ and alpha-1-acid glycoprotein) have been localized to the 9q34 region (Chan et al., 1994; Salier et al., 1992). Similarly, clusters of lipocalin genes are found in the mouse. One such cluster, containing the alpha-1-acid glycoprotein, MUP, and alpha-1-microglobulin genes has been assigned to mouse chromosome region 4B→C (Chan et al., 1994). A second group of lipocalin genes (complement component 8γ, placental protein 14, prostaglandin D2 synthase and neutrophil gelatinase associated lipocalin) has been mapped to mouse chromosome 2 by linkage analysis. The complement component 8γ, prostaglandin D2 synthase, and neutrophil gelatinase associated lipocalin have been assigned to 2B→C1 as a result of their tight linkage to the gene coding for the immunoglobulin heavy chain binding protein (Haas et al., 1992; Chan et al., 1994). The gene for placental protein 14 has been assigned to 2F3→G on the basis of its linkage with the

interleukin-1β gene (Chan et al., 1994).

The genes for alpha-1-acid glycoprotein and alpha-2u-globulin are known to be located on rat chromosome 5q (Kurtz, 1981; Szpirer et al., 1990) which is known to share syntenies with both mouse chromosome 4 and human chromosome 9.

The conserved syntenies in these three species suggests that the ancestral gene for the lipocalin superfamily may have arisen at the chromosomal region from which human 9q34, mouse 4B→C/2F3→G, and rat 5q originate (Salier et al., 1992). A large series of duplications in this ancestral region are proposed to have occurred prior to the divergence of these three species (Salier et al., 1992; Chan et al., 1994). This expansion led to the existing cluster of lipocalin genes found within conserved syntenies in these species.

The finding that lipocalin genes are divided over at least two chromosomal regions in the mouse indicates a chromosomal rearrangement with breakpoints within the lipocalin cluster occurred in the mouse after divergence from humans, and that the human arrangement has remained closer to that in the common ancestor. This is consistent with the evidence that the overall linkage map in rodents has been divided more than in primates, when compared to an ancestral mammalian genome (Lundin, 1993). The paucity of linkage analysis, and *in situ* hybridization mapping results from the rat has left the status of the chromosomal arrangement of the various lipocalin genes in this species unresolved at present. Of the 110 loci mapped to rat chromosome 5, only alpha-2u-globulin and alpha-1-acid glycoprotein are lipocalins (RATMAP). No lipocalins have yet been assigned to rat chromosome 3, the homologue to mouse chromosome 2 (RATMAP).

Regional duplications have undoubtedly played a role in the evolution of the lipocalin superfamily, however, the low degree of amino acid homology between many lipocalins, and their occurrence in arthropods and some bacterial species, indicates a very ancient origin for the family. This implies that the expansion of the lipocalin genes has also resulted from tetraploidization events in the lines of descent leading from the primitive chordates to mammals (Lundin 1993; Chan et al., 1994).

Two lipocalin genes that have been mapped to a cytogenetic location in humans do not localize to the main cluster at 9q34. The human retinol-binding protein has been mapped to 10q23→24 and the mouse homologue maps to 19D1→D2 (Chainani et al., 1991). The human apolipoprotein D gene has been mapped to 3p14→qter (Drayna et al., 1987).

It is interesting to note that these two lipocalins group more closely with arthropod and insect lipocalins than they do with any vertebrate lipocalins in a phylogenetic tree prepared from 25 lipocalin protein sequences (Igarashi et al., 1992; Figure 1.2). This suggests that an early rearrangement event separated at least these two lipocalins from the major cluster that resides at 9q34 in humans (Chan et al., 1994).

1.4. RAT ALPHA-2U-GLOBULIN

1.4.1. INTRODUCTION

Rat alpha-2u-globulin is a small (18.7 kDa, 162 amino acid) protein that was first described by Roy and Neuhaus (1966a). It is an excretory protein of mature male rat urine, but completely absent from the urine of immature male rats and female rats of all ages. Alpha-2u-globulin accounts for approximately half of the urinary protein in mature male rats, with 25-30 mg excreted daily (Roy, 1979). Alpha-2u-globulin was originally thought to be secreted from the prostate gland, however, liver perfusion, immuno-fluorescence, and *in vivo* labeling studies have demonstrated that alpha-2u-globulin is synthesized in the parenchymal cells of the liver, where it accounts for approximately 1% of total hepatic mRNA, and secreted into the serum (Roy and Neuhaus, 1966b; Roy and Raber, 1972; Kurtz et al., 1976; Sippel et al., 1976). With a molecular weight of less than 20 kDa, alpha-2u-globulin is readily filtered across the glomerulus. Approximately 60% of the glomerular filtered alpha-2u-globulin is reabsorbed back into the proximal tubules of the kidney, while the remaining 40% is excreted in the urine (Neuhaus et al., 1981). Alpha-2u-globulin is reabsorbed by endocytosis, the endosome is fused with a lysosome to form a phagolysosome (secondary lysosome) and the protein is catabolized. Alpha-2u-globulin is catabolized relatively slowly, with a half-life of 5-8 hours (Lehman-McKeeman et al., 1990).

The absence of alpha-2u-globulin from the urine of immature males, senescent males, and females of all developmental stages immediately suggested a role for hormones in modulating the expression of alpha-2u-globulin. Hypophysectomy, the removal of the pituitary gland, which is known to cause a multiple hormonal deficiency, results in the complete inhibition of hepatic alpha-2u-globulin synthesis (Kumar et al., 1969). This deficiency is not reversible by treatment with androgens. However, alpha-2u-globulin can be induced in castrated female rats treated with androgen (Roy and Neuhaus, 1967; Roy et al., 1975; Roy, 1977). These observations suggested that hepatic alpha-2u-globulin was subject to complex multihormonal regulation. This hypothesis was strengthened by the findings that adrenalectomy and thyroidectomy resulted in more than a 90% reduction

in the urinary output of alpha-2u-globulin (Irwin et al., 1971; Roy, 1973). Pharmacological and surgical experimentation has indicated the requisite involvement of at least five hormones in the hepatic synthesis of alpha-2u-globulin. Androgens, glucocorticoids, thyroid hormone, growth hormone and insulin are required to maintain normal levels of hepatic alpha-2u-globulin in male rats, while estrogens repress synthesis. The hormonal regulation of hepatic alpha-2u-globulin has been extensively studied (see 1.4.2).

Unlike the hepatic synthesis of alpha-2u-globulin, which is strictly sex-specific, several other tissues have been found to express alpha-2u-globulin genes in both male and female animals. These tissues include salivary, lachrymal, preputial, perianal, and meibomian glands of both sexes, and the mammary gland during mid-pregnancy (Laperche et al., 1983; Gubits et al., 1984; MacInnes et al., 1986; Mancini et al., 1989).

1.4.2. HORMONAL REGULATION OF HEPATIC ALPHA-2U-GLOBULIN

1.4.2.1. THYROID HORMONE

Roy (1973) reported that thyroidectomy of mature male rats depressed urinary output of alpha-2u-globulin. Kurtz et al. (1976) demonstrated, using an *in vitro* translation system, that the absence of alpha-2u-globulin in thyroidectomized males results from the lack of translatable alpha-2u-globulin mRNA. These results were confirmed by Northern blot hybridization experiments, suggesting that thyroid hormone acts pre-translationally in regulating hepatic alpha-2u-globulin synthesis (Chatterjee et al., 1983). Thyroidectomized males given [^{3}H] leucine *in vivo* show no incorporation of the label into alpha-2u-globulin, but following triiodothyronine or thyroxine treatment, alpha-2u-globulin is detected 4 days later and control levels of alpha-2u-globulin are reached within ten days (Feigelson and Kurtz 1978). The increased protein levels are paralleled by a concomitant increase in mRNA levels.

It has been difficult to demonstrate *in vivo* that the thyroid hormones are acting directly on the liver to induce alpha-2u-globulin synthesis, since thyroid hormones are known to have significant effects on circulating levels of androgens and growth hormone, both of which influence alpha-2u-globulin synthesis. The possibility that the observed effects in thyroidectomized male rats was the result of a decrease in androgen production following thyroidectomy, and a subsequent reinduction of androgens by thyroid hormone administration was examined by Kurtz et al. (1976). Treatment of thyroidectomized males with a variety of androgens failed to stimulate any detectable levels of alpha-2u-globulin. In similar experiments, thyroidectomized rats treated with growth hormone produced essentially the

same level of hepatic alpha-2u-globulin mRNA observed in intact males (Chatterjee et al.,1983). However, upon determination of alpha-2u-globulin protein concentrations in the liver cytosol of these animals it was found to be less than half that of control animals. This is in contrast to the restoration of both alpha-2u-globulin mRNA and protein levels in hypothyroid rats treated with thyroxine (Feigelson and Kurtz 1978; Chatterjee et al., 1983). This observation indicated that growth hormone may be required for maintaining the hepatic concentration of alpha-2u-globulin mRNA, while thyroxine may act at a step downstream of mRNA synthesis. This possibility was explored further by examining hepatic alpha-2u-globulin mRNA and protein levels in thyroidectomized males pretreated with growth hormone prior to the administration of thyroxine (Chatterjee et al., 1983). Results showed that there was no increase in the hepatic concentration of alpha-2u-mRNA over that observed after the pretreatment with growth hormone, but prior to thyroxine administration. However, there was an almost linear increase in the level of hepatic alpha-2u-globulin, reaching a concentration after seven days that was two fold greater than the concentration at day one of thyroxine treatment.

It seems clear from the results of these experiments that the thyroxine mediated increase in hepatic alpha-2u-globulin mRNAs in thyroidectomized rats is brought about indirectly by an increase in the circulating levels of growth hormone. The indirect nature of its action assists in accounting for the long delay (4 days, see above) observed for the thyroxine dependent stimulation of detectable levels of alpha-2u-globulin in thyroidectomized males. However, it is also apparent that thyroxine also exerts a direct effect on the utilization of hepatic alpha-2u-globulin mRNAs, since growth hormone treated thyroidectomized rats only produce half the alpha-2u-globulin observed in control animals despite their normal mRNA concentration. Similar observations have been made for the hepatic secretion of albumin in thyroidectomized male rats (Peavy et al., 1981), suggesting a role for thyroxine in the post-transcriptional regulation of a variety of proteins.

1.4.2.2. GROWTH HORMONE

Hypophysectomy of mature male rats results in the cessation of hepatic alpha-2u-globulin synthesis (Kumar et al., 1969). Reversal of the effects of hypophysectomy on alpha-2u-globulin synthesis requires the simultaneous administration of androgen, glucocorticoids, thyroxine, and growth hormone (Roy 1973).

Roy and Dowbenko (1977) and Roy et al. (1982) have investigated the concentration of hepatic and urinary alpha-2u-globulin and hepatic mRNA using: liquid hybridization kinetics, *in vitro* translation and radioimmuno-assays on samples obtained from hypophysectomized male rats

administered a multiple hormone supplementation regime that either included growth hormone, or was deficient for the peptide hormone. Similar studies carried out on hypophysectomized prepubescent rats indicated that the response to the hormone supplementation regime including growth hormone was only 10% of that observed in rats hypophysectomized as adults (Lynch et al., 1982). This observation suggests that additional pituitary factors are required during puberty to potentiate normal alpha-2u-globulin expression. The combined results of these studies provide strong evidence that growth hormone modulates the concentration of hepatic alpha-2u-globulin mRNA.

The sexually dimorphic expression pattern exhibited by hepatic alpha-2u-globulin is not unique. In the rat and mouse a number of genes with specific or preferential expression in the liver display this sexual dimorphism (Husman et al., 1985; Johnson et al., 1995). The observation that growth hormone displays a sexually dimorphic secretory rhythm in rats (Eden 1979), prompted the investigation of a link to the observed differences in hepatic alpha-2u-globulin expression in male and female rats (Husman et al., 1985). Growth hormone is secreted in a continuous fashion in females, while secretion in males is periodic. The continuous infusion of growth hormone (along with corticosterone, thyroxine and dihydroxy-testosterone) in hypophysectomized males did not restore urinary output of alpha-2u-globulin, whereas the same hormone replacement regime administered as a daily injection restores alpha-2u-globulin production to near control levels (Husman et al., 1985). The absence of alpha-2u-globulin in the urine of female rats may, then, very well be related to the more or less continuous release of growth hormone in this sex.

Extensive DNAse I footprinting analysis of the alpha-2u-globulin promoter and gene sequences have been carried out by a number of groups (Addison and Kurtz 1989; Choy et al., 1989; Sierra et al., 1990; Van Dijck et al., 1993a). One identified footprint (footprint B, Sierra et al., 1990; Van Dijck et al., 1993a) in the promoter proximal region shares similarity with a growth hormone response element identified in the promoter region of the SPI 2.1 gene (Yoon et al., 1990; Paquereau et al., 1992; Enberg et al., 1994). The nuclear factor(s) responsible for the footprint have yet to be identified. A similar element, bearing homology to the growth hormone response element, has been identified in the mouse MUP gene, which is the mouse homologue of alpha-2u-globulin (discussed below) (Johnson et al., 1995). The activity of the SPI 2.1 growth hormone response element is dependent on the presence of a nearby C/EBP binding site. A C/EBP binding site has been identified in the promoter proximal region of the alpha-2u-globulin and MUP genes (Sierra et al., 1990; Van Dijck et al., 1993a; Johnson et al., 1995).

Deletion of a region including the putative growth hormone response element has been correlated with reduced expression of a hybrid MUP

transgene (Johnson et al., 1995). However, no difference in protein binding to this region has been observed between extracts from mature males, prepubertal males, mature females and hypophysectomized males (Van Dijck et al., 1993a). Furthermore, templates containing the putative growth hormone response element are not differentially transcribed *in vitro* using nuclear extracts derived from male and female rats (Sierra et al., 1990).

The lack of sexual dimorphism in the studies reported above may indicate the involvement of additional elements not present in the templates examined. Johnson et al. (1995) examined the possibility that elements within introns of the MUP gene contribute to the correct expression pattern. Transgenic lines prepared from a construct lacking intron 1 demonstrated a significant decrease in expression of the transgene. Van Dijck et al. (1993b) identified two DNAse I footprints in the first intron of alpha-2u-globulin. One site is a proposed binding site for the liver specific transcription factor pseudo NF-1, while the other site has been characterized as a functional AP1 site binding the c-Fos/c-Jun complex.

The potential significance of sequences downstream of the transcriptional start site of alpha-2u-globulin was demonstrated with *in vitro* transcription assays on an alpha-2u-globulin template that contained sequences from -639 to +1395 (Sarkar and Feigelson, 1989). In contrast to the in vitro transcription studies carried out by Sierra et al. (1990), the template utilized in this study, which includes the first three introns, was transcribed only by nuclear extracts derived from mature male rat livers.

Members of the Jun and Fos families are known to be induced in osteoblasts and preadipocytes when they are stimulated by growth hormone (Slootweg et al., 1990, 1991; Doglio et al., 1989; Gurland et al., 1990). Given these observations a mechanism for the growth hormone stimulation of MUP and alpha-2u-globulin has been proposed (Johnson et al., 1995). These researchers suggest that during continuous growth hormone stimulation, as is the case in females, the Fos/Jun complex would be down-regulated, leading to a low level of hepatic transcription of MUP and alpha-2u-globulin in females. This is analogous to the reduced levels of Fos and Jun observed during the continuous activation of protein kinase C (Slootweg et al., 1991). Conversely, when growth hormone stimulation is intermittent, as it is in males, the appropriate members of the Fos and Jun families can be successively activated by the pulses of growth hormone leading to the activation of hepatic MUP and alpha-2u-globulin.

1.4.2.3. INSULIN

The daily urinary output of alpha-2u-globulin is drastically reduced in diabetic rats, however, this defect is reversible through the administration of

insulin (Roy and Leonard, 1973). A large discrepancy between urinary output and hepatic mRNA levels is observed in diabetic animals. Additionally, there is a lack of correlation between hepatic mRNA and protein concentrations indicating that insulin may be involved at both a transcriptional and post transcriptional level of alpha-2u-globulin expression (Roy, 1979).

Using a rat hepatoma cell line Widman and Chasin (1982) examined the induction of alpha-2u-globulin by insulin. The induction profile was complex. Alpha-2u-globulin synthesis was shown to increase dramatically at insulin concentrations expected to saturate the insulin receptors. At increased, but still physiologically relevant, insulin concentrations a gradual decrease in alpha-2u-globulin synthesis was observed. This is suggestive of both a positive and negative regulatory effect for insulin on alpha-2u-globulin synthesis. At elevated, non-physiological, insulin concentrations the authors once again observed an increased rate of alpha-2u-globulin synthesis. This induction at non-physiological concentrations of insulin was surmised to be the result of insulin cross-reacting with other hormone receptors. Such cross reactivity has been documented for insulin and the somatomedin receptor (Van Wyck et al., 1975; Takano et al., 1975). A similar induction of alpha-2u-globulin at high concentrations of insulin was observed in hypophysectomized rats (Kurtz et al., 1978a).

1.4.2.4. SEX HORMONES

The observed sexually dimorphic expression of hepatic alpha-2u-globulin has lead to numerous investigations into the role sex hormones play in regulating expression of this gene. Alpha-2u-globulin synthesis has been shown to decline by approximately 90% in males castrated as adults, and is totally absent in prepubescent castrated males (Kurtz et al., 1976). Administration of dihydrotestosterone to males castrated as adults results in the restoration of urinary and hepatic concentrations of alpha-2u-globulin (Kurtz et al., 1976). However, in castrated prepubescent males there is no response to androgen treatment (Roy, 1979). This is likely the result of the lack of androgen sensitivity in liver tissue prior to 40 days of age (Roy, 1972). The treatment of castrated females with dihyrotestosterone results in the induction of hepatic alpha-2u-globulin synthesis in females. The increase in alpha-2u-globulin concentration in both the treated castrated male and female animals is paralleled by an increase in the functional mRNA levels as determined by *in vitro* translation analysis, suggesting that androgens regulate alpha-2u-globulin synthesis pretranslationally and possibly at the level of transcription (Kurtz et al., 1976). The *in vitro* binding of androgen receptor complexes with sequences in the 5′ upstream region of an alpha-2u-globulin gene is consistent with a mechanism in which androgens exert their regulatory effect directly (Van Dijck et al., 1987).

Treatment of intact males with estrogens gradually depresses the synthesis of alpha-2u-globulin (Kurtz et al., 1976). After 8 days of estrogen treatment alpha-2u-globulin is undetectable in the liver and urine, and a parallel decrease in functional mRNA for alpha-2u-globulin is evident (Kurtz et al., 1976). The reversal of these effects is not immediate following cessation of the estrogen treatment, and the administration of androgens during the lag phase does not hasten the induction of alpha-2u-globulin synthesis (Roy, 1979)

It is unclear if estrogens act directly in the liver to modulate alpha-2u-globulin expression. Estrogens have been suggested to inhibit the synthesis of testosterone from its precursors in testicular tissues, and to decrease the circulating concentrations of luteinizing hormone, secreted from the pituitary, which is required for testosterone synthesis (Feigelson and Kurtz, 1978). Roy et al. (1974) reported that estrogens interfere with hepatic uptake of androgens possibly through reducing expression of, or otherwise inactivating, their cognate receptors. Furthermore, estrogens are known to exert an effect on the secretory pattern of growth hormone, generating a characteristically female pattern of secretion in males treated with estrogens (Mode et al., 1982). Any of these mechanisms could account for the decreased biosynthesis of alpha-2u-globulin after estrogen treatment.

Recently (Van Dijck and Verhoevan, 1992), demonstrated the *in vitro* binding of the estrogen receptor to two regions in the 5′ upstream region of an alpha-2u-globulin gene. One of the sites contains an imperfect palindromic sequence resembling the consensus estrogen response element. The other site is a perfect estrogen response element half site. The binding affinity for these sites however, is significantly lower than the affinity estrogen displays for an oligonucleotide containing a canonical response element. The physiological significance of these binding sites in the estrogen mediated repression of alpha-2u-globulin, if any, has not been addressed, however, a clear footprint over the imperfect palindromic element has been observed with liver nuclear extracts (Van Dijck and Verhoevan, 1992).

1.4.2.5. GLUCOCORTICOIDS

Of the many classes of hormones that regulate alpha-2u-globulin expression, the regulation by glucocorticoids is the best understood. Castration of adult male rats results in the reduction of both the synthesis of hepatic alpha-2u-globulin and its mRNA (Kurtz et al., 1976). Unlike the slow 8-10 day androgen treatment required to restore alpha-2u-globulin expression, administration of glucocorticoids to castrated males brings about an increase in alpha-2u-globulin synthesis within 2-3 hours with control levels reached within 24 hours (Kurtz et al., 1978b). The increase in alpha-2u-globulin is paralleled by an increase in mRNA indicating a pretranslational

control mechanism (Kurtz et al., 1978b). This rapid reinduction requires the glucocorticoid receptor, so cannot be attributed to the metabolic conversion of glucocorticoids to androgens (Kurtz et al., 1978b). Similar results were obtained in studies on glucocorticoid stimulated hepatocyte suspensions isolated from castrated males (Chen and Feigelson, 1978a; 1978b).

Studies with a permanent rat hepatoma cell line, and mouse L cells transfected with alpha-2u-globulin genes have indicated that the response to glucocorticoids is reproducible in permanently established cell lines (Kurtz 1981b; Widman and Chasin, 1982; Addison and Kurtz, 1986). In both these cases the induction curve is biphasic, with an initial lag phase of 2-4 days preceding a period of steadily increasing rates of alpha-2u-globulin synthesis before steady state levels are attained after about 9 days (Widman and Chasin, 1982; Addison and Kurtz, 1986). The long lag period is consistent with the idea that glucocorticoids act indirectly in regulating alpha-2u-globulin synthesis, requiring the accumulation of one or more intermediate gene products to mediate its effects. This notion is supported by the observations that alpha-2u-globulin induction by glucocorticoids is dependent on protein synthesis during the lag phase in both primary hepatocyte suspensions and permanent cell lines (Chen and Feigelson, 1979; Addison and Kurtz, 1986).

Transfection assays have indicated that the response to glucocorticoids is retained even after trimming the 5′ upstream region to 235 bp and the 3′ flanking sequences to 400 bp (Kurtz et al., 1982). Extensive linker-scanning mutagenesis of the 235 bp upstream sequences has identified two regions involved in the induction by glucocorticoids (Addison and Kurtz, 1986). Mutations between -115 and -160 defined the regulatory region, and abolished, or greatly reduced alpha-2u-globulin induction. The regulatory region displays dyad symmetry. A 6 bp sequence is separated from its inverted complement by 22 bp. The 6 bp sequences are contained within a 12 bp stretch which displays a reduced degree of dyad symmetry. The promoter proximal element of dyad symmetry (-116 to -137) is absolutely conserved among 14 alpha-2u-globulin promoters examined in this region (Addison and Kurtz, 1986). The remainder of the regulatory region is highly conserved among the same 14 alpha-2u-globulin promoters. Interestingly, this region also exhibits marked homology (90%) to the mouse MUP genes in this region (Addison and Kurtz, 1986). Surprisingly, mutations between -177 and -200 were found to increase induction 2-4 fold.

DNase I footprinting of the region between -110 and -220, which encompasses the glucocorticoid regulatory region and the hyperinducibilty region, has revealed four footprints (Addison and Kurtz, 1989; Sierra et al., 1990; Van Dijck et al., 1993a). The two most promoter-proximal footprints map to the regulatory region. Each footprint protects one of the 6 bp inverted repeats present in the region of dyad symmetry. The sequence protected,

GCCAAG, is a binding site for the liver-enriched transcription factor pseudo NF-1 (Sierra et al., 1990; Van Dijck et al., 1993a). The remaining two footprints are localized to the hyperinducibility region. The most distal footprint covers a binding site for the liver-specific transcription factor C/EBP (Sierra et al., 1990; Van Dijck et al., 1993a). Since both pseudo NF-1 and C/EBP are constitutively expressed in hepatic tissue, they are clearly not the regulatory factors synthesized in response to glucocorticoids that bring about the induction of alpha-2u-globulin. Addison and Kurtz (1989) suggest three simple models for the induction of alpha-2u-globulin by glucocorticoids that are consistent with the footprinting data. Firstly, pseudo NF-1 may be bound to the two elements in the regulatory region at all times, but require some modification in response to glucocorticoid treatment before it can activate the promoter. Secondly, bound pseudo NF-1 may provide the sequence specificity, while another protein, synthesized in response to hormone treatment, could bind to the pseudo NF-1:DNA complex to stimulate transcription. Finally, access to the pseudo NF-1 binding sites may be regulated by glucocorticoid treatment of the cell.

Van Dijck et al. (1987) identified several glucocorticoid receptor binding sites in the 5′ flanking sequences of an alpha-2u-globulin using *in vitro* binding assays. The functional significance of these sites is unclear at present. Several of the identified sites are outside the 235 bp region capable of directing the glucocorticoid mediated induction of alpha-2u-globulin in transfected cells. Deletion of the putative receptor binding sites has no effect on the induction of alpha-2u-globulin in transfected cell lines (Addison and Kurtz unpublished results).

Hess et al. (1990) identified a 206 bp fragment from the fourth intron of alpha-2u-globulin which contained multiple *in vitro* glucocorticoid receptor binding sites. The receptor binding sites are organized into at least three independent binding clusters (Chan et al., 1991). The 206 bp fragment and numerous sub-fragments were functionally characterized in stable transfection assays, and were shown to confer secondary glucocorticoid responsiveness on a thymidine kinase promoter driving chloramphenicol acetyltransferase (CAT) expression. A lag of about 10 hours was observed, and the induction of CAT mRNA was inhibited by cyclohexamide. The redundancy of receptor binding sites appears to be important for attaining maximal induction, but has no influence on the duration of the lag period (Chan et al., 1991). Mutational analysis of these putative secondary glucocorticoid hormone response elements in the context of the alpha-2u-globulin promoter and gene sequences coupled with transfection assays will be required to fully resolve their physiological significance in glucocorticoid induction of alpha-2u-globulin.

1.4.3. NON-HEPATIC ALPHA-2U-GLOBULIN

A number of tissues other than liver express alpha-2u-globulin in both male and female rats. These tissues are secretory in nature and include preputial, lachrymal, meibomian, submaxillary, perianal and mammary glands (Gubits et al., 1984; MacInnes et al., 1986; Mancini et al., 1989).

Immuno-localization of alpha-2u-globulin and *in situ* hybridization with an alpha-2u-globulin cDNA indicated that alpha-2u-globulin and its mRNA are concentrated in the acinar cells of the lachrymal gland (Gubits et al., 1984). Developmental dot blot analysis of lachrymal mRNA isolated from males and females indicates that lachrymal alpha-2u-globulin mRNA is detectable as early as 6 days postpartum (Gubits et al., 1984). This contrasts with the hepatic expression of alpha-2u-globulin which is not observed until the onset of puberty at about 30 days post partum (Roy et al., 1974). Lachrymal gland mRNA gradually increases in males until day 30 when adult levels are attained and maintained (Gubits et al., 1984). A somewhat more complex increase is observed in females (Gubits et al., 1984). A sexual dimorphic expression pattern is seen for lachrymally expressed alpha-2u-globulin, with a 3-5 fold higher level of expression observed in males (MacInnes et al., 1986). Hypophysectomized males and females show about a 90% decrease in lachrymal alpha-2u-globulin-mRNA levels, indicating that expression in this tissue is under hormonal control (Gubits et al., 1984). The hormonal requirements for expression in the lachrymal gland are less stringent than that observed for hepatic expression as determined by dot blot analysis of RNA isolated from hypophysectomized animals treated with various hormonal replacement regimes (Gubits et al., 1984). Isoelectric focusing of in vitro translated mRNAs revealed that an acidic subset of about 5 alpha-2u-globulins are expressed in the lachrymal gland (MacInnes et al., 1986). Quantitative mRNA dot blot analysis indicates that female and male lachrymal glands express, respectively, levels 10% and 30% of that observed in adult male livers (MacInnes et al., 1986).

The submaxillary gland expresses a subset of acidic alpha-2u-globulin isoforms that is almost identical to that observed for the lachrymal gland, but clearly distinguishable from the more basic, hepatically expressed isoforms (Laperche et al., 1983; Gubits et al., 1984; MacInnes et al., 1986). Alpha-2u-globulin was immunolocalized to the striated and intercalated ductal cells of the submaxillary gland (Gubits et al., 1984). Submaxillary gland alpha-2u-globulin mRNA does not display a sexually dimorphic expression pattern, and is unaffected by hypophysectomy (Laperche et al., 1983; Gubits et al., 1984). Submaxillary gland transcripts are detected as early as 6 days postpartum in both males and females (Gubits et al., 1984; MacInnes et al., 1986). Expression is highest prior to weaning and decrease abruptly following puberty (Laperche et al., 1983; Gubits et al., 1984; MacInnes et al., 1986). A

diverse set of salivary gland mRNAs for alpha-2u-globulin have been observed. The alternative usage of at least six splice acceptor sites located in the sixth intron and untranslated seventh exon regions, coupled with three polyadenylation sites, produces a set of mRNAs with diverse 3′ non-coding regions (Gao et al., 1989). Submaxillary transcript levels from 30 day-old males are approximately 25% that of transcripts from a mature male liver (Laperche et al., 1983; MacInnes et al., 1986).

Eighteen isoforms of alpha-2u-globulin have been detected in the preputial glands of both male and females (MacInnes et al., 1986). Preputial alpha-2u-globulin has been immuno-localized to the acinar cells of the preputial gland (Mancini et al., 1989). A subset of the expressed isoforms is observable by Western blot analysis as early as 10 days postpartum, with the entire complement detected by day 25 (MacInnes et al., 1986). Synthesis of alpha-2u-globulin peaks at day 35-40, and drops slightly in adulthood (MacInnes et al., 1986). With the exception of the two most acidic lachrymal gland isoforms, all isoforms expressed in other tissues are detectable in the preputial gland (MacInnes et al., 1986). Since the number of detectable alpha-2u-globulin isoforms in the preputial gland is close to the estimated number of alpha-2u-globulin genes in the genome, it is possible that the entire complement of genes is transcriptionally active in this tissue. It is not surprising then, that the preputial gland exhibits the highest level of alpha-2u-globulin expression of any tissue. Alpha-2u-globulin transcripts represent 5-10% of all preputial gland mRNA, and reaches levels approximately 3-4 fold higher than those observed in the adult male liver (MacInnes et al., 1986).

Expression of alpha-2u-globulin is first detectable in the mammary glands of pregnant females at about day 15 of pregnancy, peaking at days 18-19 and becomes undetectable by day 21 (MacInnes et al., 1986). Alpha-2u-globulin is not detectable in nursing animals, or after pups were weaned (MacInnes et al., 1986). The same 18 isoforms present in the preputial gland are observed in the mammary glands (MacInnes et al., 1986). Compared to all other expressing tissues, there are relatively low levels of alpha-2u-globulin mRNA in the mammary glands of pregnant females. Levels of mRNA for the mammary glands are about 5% of that of the adult male liver (MacInnes et al., 1986).

Little is known with respect to expression in the perianal and meibomian glands, the remaining two tissues known to express alpha-2u-globulin. *In situ* hybridization to these two tissues has demonstrated that low levels of alpha-2u-globulin transcripts are localized to select perenchymal epithelial cells and hair shaft keratinocytes and accessory oil glands (Mancini et al., 1989).

1.4.4. ALPHA-2U-GLOBULIN AND MUP GENOMIC STRUCTURE

Genomic Southern analysis of rat DNA with an alpha-2u-globulin cDNA probe reveals a complex pattern of bands consistent with a multigene family (Kurtz 1981). Although the pattern of bands is complex, it appears identical among DNA isolated from several tissues of individual male and female rats, as well as from several established cell lines (Kurtz 1981; Dolan et al., 1983). The band intensity observed on genomic Southerns are nonmolar, i.e., there are select bands in each genomic digest that are very much more intensely hybridizing than the others, indicating that many of the family members share several restriction sites, while a smaller number of family members are divergent for these sites. The number of alpha-2u-globulin genes has been estimated from solution hybridization kinetics to be approximately 18-20 (Kurtz 1981). A similar estimate of gene number has been arrived at following isoelectric focusing electrophoresis of the protein isoforms generated from *in vitro* translation of mRNA from expressing tissues (MacInnes et al., 1986; Laperche et al., 1983).

Since posttranslational processing does not occur in the reticulocyte lysate *in vitro* translation system the numerous isoelectric forms observed are thought to reflect different mRNA sequences. These mRNAs could arise as a result of alternative splicing of the same gene transcript, or by transcription from numerous different genes. The later hypothesis is favored since the observed alternative splicing patterns (see below) appear to be confined to the 3′ untranslated region.

Characterization of nucleotide sequences coupled with electron microscope analysis of RNA-DNA heteroduplex indicates that the alpha-2u-globulin gene is made up of 7 exons and encompasses about 4kb of genomic DNA, from the transcription initiation site to the polyadenylation sites (Figure 1.3) (Kurtz 1981; Dolan et al., 1982, 1983). Furthermore, sequencing has shown that the seventh exon is entirely non-coding, as the translational termination signal is within the sixth exon (Dolan et al., 1982). The first exon codes for a 19 amino acid hydrophobic signal sequence responsible for secretion, as well as the first 14 amino acids of the mature protein (Drickamer et al., 1981). Sequence analysis has also revealed that both alpha-2u-globulin and MUP belong to the lipocalin superfamily of proteins (reviewed in section 1.4) and as such are likely to be involved in the binding and/or transport of small hydrophobic molecules involved in signal transduction (Pervais and Brew 1985).

The size of the major transcript in three alpha-2u-globulin expressing tissues (lachrymal gland, salivary gland, and liver) is the same (approximately 1.2 kb) (Dolan et al., 1982; Gubits et al., 1984; Laperche et al., 1983; Gao et al., 1989). However, in each of these tissues there are

minor transcripts present, indicative of alternative splice variants.

Sequence comparison of a salivary and hepatic cDNA indicated that the splice junctions were identical within the coding regions (Laperche et al., 1983). However, splice-acceptor site heterogeneity in the removal of the sixth intron has been detected in both liver and salivary gland (Laperche et al., 1983; Gao et al., 1989). Among 30 salivary gland cDNAs Gao et al. (1989) found evidence of the alternative usage of 6 splice acceptor sites in the sixth intron-seventh exon regions. The functional significance of alternative splicing in the sixth intron-seventh exon regions is unclear given the fact that the length heterogeneity generated is entirely within the untranslated region of the transcript. One possibility is that these alternatively spliced transcripts may vary in terms of their stability, however this is speculative and has not been directly investigated. The retention of an entirely non coding seventh exon may indicate a role for these sequences in the regulation of alpha-2u-globulin expression.

It is interesting to note that in the liver two of the minor transcripts are larger than the gene with sizes of 4.2, and 6.6 kb (Dolan et al., 1982). The 6.6 kb liver transcript hybridizes with a probe derived from the 5' flanking sequences of alpha-2u-globulin, resembling the immature transcripts of the human U1 RNAs which are extended upstream of the putative CAP sites (Manser and Gesteland, 1982), and may then represent an immature transcript.

In addition to the alternative splicing patterns in the 3' region of the gene, the differential usage of at least three polyadenylation sites (Dolan et al., 1983; Gao et al., 1989) adds further to the diversity of mRNAs transcribed from the alpha-2u-globulin genes.

Like alpha-2u-globulin, the MUPs are a group of closely related proteins which are synthesized in the liver of mice in significant quantities and excreted in the urine (Finlayson et al., 1965). The hepatic expression of MUP is sexually dimorphic, with males synthesizing up to 20 fold more MUP (Finlayson et al., 1965). The hepatic expression of the MUP genes is under multihormonal control by thyroid hormone, growth hormone, insulin, glucocorticoids and androgens (Hastie et al., 1979; Shaw et al., 1983; Knopf et al., 1983). MUP mRNA comprises about 5% of all hepatic mRNA (Hastie and Held 1978). Expression of MUPs has also been detected in lachrymal, submaxillary and mammary glands (Hastie et al., 1979; Shaw et al., 1983). Hormones appear to have no effect on the synthesis of MUPs in the submaxillary gland, whereas testosterone has been identified as a regulator of expression in the lachrymal gland (Knopf et al., 1983).

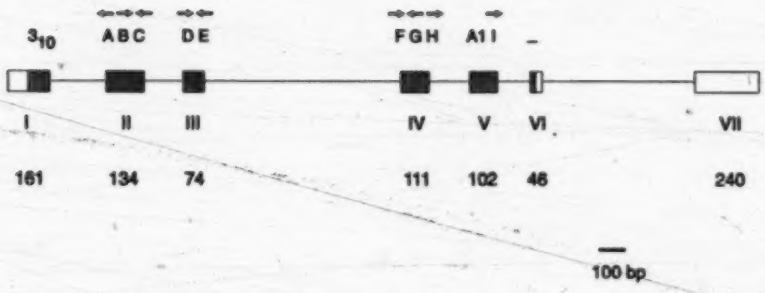


Figure 1.3. Structure of the alpha-2u-globulin gene. Exons are indicated by the boxes, with the shaded regions indicating the coding portions. Introns are indicated by the lines. Below the figure the exons are numbered I-VII, and the number of nucleotides in each exon are indicated. Above the figure the location of β-sheets A-I as indicated in Figure 1.1 are shown, their antiparallel nature is indicated by the arrows. The N-terminal and C-terminal coils are indicated by 3_{10} and the _ respectively. The position of the α-helix is labeled A1. The location of these three-dimensional elements is for reference only and are not to scale.

Bishop et al. (1982) estimated the number of MUP genes to be approximately 35. Furthermore, on the basis of cross hybridization, the MUP genes were placed into two major groups, each containing about 15 members, with a small number of genes not belonging to either group. Most if not all of the MUP genes were localized to mouse chromosome 4 by somatic cell hybrid panel screening (Bishop et al., 1982). The Group 1 genes differ from each other by about 2%, while the Group 2 genes are slightly more divergent (Clark et al., 1984b). The Group 1 and Group 2 genes differ from each other by about 10% (Clark et al., 1984b). The Group 1 genes appear to be fully functional active MUP genes, whereas analysis has indicated that the Group 2 genes contain lesions that would result in their classification as pseudogenes (Ghazal et al., 1985; Clark et al., 1985a; Clark et al., 1985b, Shahan et al., 1987b). The Group 1 genes are expressed in the liver (Clark et al., 1985a; Shahan et al., 1987a; Shahan et al., 1987b). The small number of genes that do not belong to either Group 1 or Group 2 appear to account for the MUP expression in the non-liver tissues (Shahan et al., 1987a; Shahan et al., 1987b).

Analysis of lambda genomic clones indicated that there was extensive homology between the clones (Clark et al., 1982). Homology between genes of both Groups 1 and 2, although sometimes interrupted, extends 4-5 kb into the 5′ flanking region and 10-12 kb into the 3′ flanking region (Bishop et al., 1982). None of these genomic clones, isolated on the basis of their hybridization to a MUP cDNA, contained more than one MUP gene, nor was homology detected between the 5′ end of one clone and the 3′ end of another clone (Bishop et al., 1982). Given that the distance between MUP genes appeared to be greater than 20 kb, Clark et al. (1984b) screened a lambda genomic library with probes derived from the 5′ flanking region of the genomic clone with the most extensive upstream sequences. A similar screen was performed using probes derived from the 3′ flanking sequences. Analysis of the clones recovered indicated that Group 1 and Group 2 genes are arranged in a head-to-head fashion, with one gene of each group present in the divergently arranged pair (Clark et al., 1984b). Clark et al. (1984b) also demonstrated the 3′ tail-to-tail linkage of adjacent sets of gene pairs.

The number of Group 1 and Group 2 genes in the mouse genome is approximately equal (Bishop et al., 1982). This is because the predominant organization of the MUP locus is an array of 45 kb units each consisting of a Group 1 gene linked to a Group 2 gene in a head-to head fashion (Figure 1.4) (Clark et al., 1984b; Bishop et al., 1985). The 45 kb units are large palindromes (Clark et al., 1984b; Bishop et al., 1985). The central region of the 45 kb palindrome is an approximately 8 kb stretch of sequences, common to many if not all of the 45 kb units, which is flanked on both sides by a 9.5 kb region of inverted symmetry which includes the 4 kb structural gene, approximately 3.5 kb of upstream sequences and 2 kb of downstream

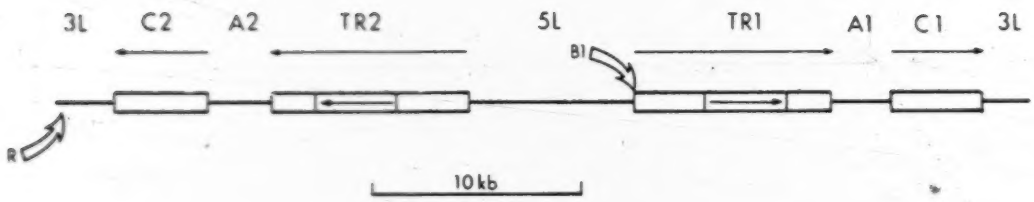


Figure 1.4. Structure of the MUP 45 kb palindrome unit. The regions of inverted symmetry are shown as boxes marked with arrows above them. TR1 and TR2, group 1 and group 2 transcription units, respectively, and their 5′ and 3′ flanking regions. The divergently transcribed group 1 and group 2 genes are shown as boxes containing arrows, which indicate the direction of transcription. Regions C1 and C2 are conserved regions of inverted symmetry found 3′ to the structural gene. Horizontal lines represent regions without inverted symmetry (A1, A2, 3L). These regions are frequently similar, or identical in different repeats. The position of two repetitive elements, R and B1, which are present in every unit examined are indicated by the open arrows. Modified from Bishop et al. 1985.

sequences (Bishop et al., 1985). A second set of sequences with inverted symmetry are found further 3′ to the structural genes (Bishop et al., 1985). Some of the 45 kb units have been found to be directly oriented, separated by 4 kb of unique sequences indicating a total separation of about 26 kb between the 3′ ends of the Group 1 and Group 2 genes (Bishop et al., 1985).

Comparisons among DNA sequences for various members of the rat alpha-2u-globulin family indicate a sequence homology of 95%-98% between members (Dolan et al., 1982; Dolan et al., 1983; Yamamoto et al., 1989). A detailed comparison of two genomic clones indicated an overall homology of 93%; ranging from 90.1%-99.0% in exonic sequences, 90.4%-93.4% in intronic sequences, 90.0% in the first 230 bp upstream of the genes and 98.3% in the 240 bp downstream of the genes (Yamamoto et al., 1989). Of the genomic clones for which sequences are available, all but one would appear to encode a functional gene (Dolan et al., 1982; Winderickx et al., 1987; Yamamoto et al., 1989). The one pseudogene identified thus far lacks an ATG start codon and possess a frame shift mutation that would result in premature termination (Winderickx et al., 1987).

Since both the rat alpha-2u-globulin and the mouse MUP genes are encoded by moderately large gene families, it seems likely that the duplication that generated these families took place before the speciation of mice and rats which is estimated to have taken place 10-30 million years ago (Clark et al., 1984a; Laird et al., 1969). The coding region of alpha-2u-globulin and MUP differ by about 20% at the nucleotide level, and approximately 30% at the amino acid level (Clark et al., 1984a; Winderickx et al., 1987). In the absence of some corrective force, one would expect the alpha-2u-globulin genes to display about a 20% divergence, however, sequence analysis indicates that the alpha-2u-globulins diverge by approximately 5% at the nucleotide level (Dolan et al., 1982), indicating that divergence has not occurred, or a homogenizing mechanism has acted on the alpha-2u-globulin gene cluster to reduce divergence. A similar phenomenon is observed for the MUP genes, which show a divergence of about 10% overall, but within the two major groups divergence between members is less than 5% (Bishop et al., 1985).

1.4.5. ALPHA-2U-GLOBULIN FUNCTION

Although the alpha-2u-globulin protein has been well characterized at several levels, its exact function has not been clearly defined. Several lines of evidence suggest that both alpha-2u-globulin and mouse MUP have an important function. Firstly, the daily urinary output of several milligrams of protein is an extreme energetic expense for animals that are often on the brink of starvation in the wild (Bacchini et al., 1992). Secondly, in both the rat and mouse the genes encoding these proteins are multicopy (Kurtz 1981;

Bishop et al., 1982). Gene duplication resulting in families with uniform members, for example, alpha-2u-globulin and the MUPs, is often the result of a demand for a large quantity of gene product (Ohta, 1991). Thirdly, there is an approximately four fold difference in the frequency of nucleotide substitutions at silent sites versus replacement sites between alpha-2u-globulin and MUP, indicating selective pressure against changes in amino acid sequence (Dolan et al., 1982; Clark et al., 1984a; Clark et al., 1984b).

The ability of both alpha-2u-globulin and MUP to bind 2-(sec-butyl) thiazoline and 2,3,-dehydro-*exo*-brevicomin *in vitro* suggests that these two proteins act as sex pheromone binding proteins (Bacchini et al., 1992). This possibility is strengthened by the observation that both alpha-2u-globulin and MUP have been purified and crystallized with bound ligands (Bocskei et al., 1992). In the case of MUP the ligand was positively identified as the pheromone 2-(sec-butyl)thiazoline, while in the case of alpha-2u-globulin the ligand was not identified, but the electron density was consistent with that of a number of known pheromones (Bocskei et al., 1992). Alpha-2u-globulin is also capable of binding many of the odorants that bind to the odorant binding proteins (Cavaggioni et al., 1990). Sex pheromone binding proteins have also been described in insect antennae (Vogt et al., 1988; Van den Berg and Ziegelberger, 1991) and boar saliva (Booth, 1984; Booth and White, 1988). A role in a function as important as sexual communication could explain the costly urinary loss of the protein, the gene amplification, and selective pressure to conserve the amino acid sequences (Bacchini et al., 1992).

Binding of pheromones to alpha-2u-globulin and MUP would allow for the concentration of the pheromones in the urine. Male rat urine is known to affect sexual responses in female rats, including sexual attraction and aggression-submission behavior (Cowley, 1978). Additionally, the odor of male urine can accelerate the attainment of puberty in the female and can block pregnancy (Vandenbergh, 1969; Parkes and Bruce, 1962). Vandenbergh et al. (1975) demonstrated that the element responsible for these activities was a protein. Pheromones may diffuse into the air in the field, following their excretion in urine. The volatility of the two pheromones shown to bind alpha-2u-globulin and MUP is fairly low, so it is more likely that the entire protein-pheromone complex reaches the vomeronasal organ through licking action (Bacchini et al., 1992). Whether or not alpha-2u-globulin and MUP also present the pheromone to specific receptors has not been established, but would be consistent with the action of other lipocalin family members.

The presence of another member of the lipocalin superfamily, the odorant binding proteins, in the nasal cavity of both mice and rats, has led to speculation that pheromone molecules may be released from their complexes with alpha-2u-globulin and MUP in the nasal cavity and subsequently bind

the similarly structured odorant binding proteins to activate a membrane receptor that initiates a signal transduction pathway (Pes and Pelosi, 1995). The use of similar proteins for the delivery and perception of a chemical message ensures that the transmitter and receiver are similarly "tuned" (Pes and Pelosi, 1995).

The proposed role for alpha-2u-globulin in the binding and stabilization of pheromones is strengthened by the fact that the preputial, meibomian and perianal glands which express alpha-2u-globulin posses holocrine secretory activity and are rich in pleiomorphic peroxisomes, which are active in the synthesis of pheromonal lipids (Kolata, 1987). More than 25% of the secreted protein from the preputial gland can be accounted for by alpha-2u-globulin (Murty et al., 1987). The preputial gland is known for its pheromonal activity, secretions from which have been demonstrated to accelerate or synchronize estrus in female rats (Chipman and Albrecht, 1974). Additionally, the perianal licking of newborn rats by their mothers has been demonstrated to serve a pheromonal function in establishing the mother-pup bond (Kolata, 1987).

Saliva is also a means through which chemical communication between animals can take place. This phenomenon has been most extensively studied in the boar (Booth and White, 1988). An extracellular protein from boar saliva, pheromaxein, which binds the known pheromone androstenol and which has strong pheromonal activity for the sow has been partially characterized (Melrose et al., 1971; Booth and White, 1988). Presently, not enough data is available to determine if boar pheromaxein is a lipocalin. It is interesting to note that alpha-2u-globulin is also expressed in the submaxillary salivary glands in the rat, thus the possibility exists that alpha-2u-globulin expressed from this tissue may also have pheromone binding activity.

The similarity between alpha-2u-globulin and aphrodisin, a lipocalin with demonstrated pheromonal binding activity, also supports a similar pheromonal binding activity for alpha-2u-globulin (Singer and Macrides, 1990).

Alpha-2u-globulin has also been implicated in a male rat-specific nephrotoxicity which develops following exposure to numerous organic chemicals (Alden et al., 1984; Alden 1986). The nephrotoxicity manifests itself, after acute exposure to certain chemicals, as the excessive accumulation of protein (hyaline) droplets in the renal proximal tubule cells (Swenberg et al., 1989; Borghoff et al., 1990). After chronic exposure to many of these agents renal tubular tumors result (Swenberg et al., 1989; Borghoff et al., 1990; Lehman-McKeeman, 1993)

It has been demonstrated, both *in vitro* and *in vivo*, that the parental chemical agents and/or their metabolites that cause hyaline droplet nephropathy and the ensuing renal tumors bind reversibly to alpha-2u-globulin (Lock et al., 1987; Lehman-McKeeman et al., 1989; Strasser et al., 1988; Charbonneau et al., 1989; Lehman-McKeeman et al., 1991), but do not affect the synthesis of alpha-2u-globulin (Olson et al., 1987). The rate of lysosomal degradation of alpha-2u-globulin bound to these agents has been extensively investigated, and found to be catabolized at a much slower rate than alpha-2u-globulin in untreated control groups (Charbonneau et al., 1988; Lehman-McKeeman et al., 1990b). The combination of the high delivery rate of alpha-2u-globulin to the proximal tubule cells of the kidney, coupled with the reduced rate of lysosomal degradation observed for alpha-2u-globulin complexed with these chemical agents results in the significant accumulation of hyaline droplets, the first step in the development of tumors (Hard et al., 1993).

With chronic exposure to hyaline inducing agents, the phagolysosomes become greatly enlarged and take on a crystalloid appearance as the result of high concentrations of pure alpha-2u-globulin (Hard et al., 1993). The accumulation of alpha-2u-globulin is cytotoxic and results in single cell necrosis, although the exact cause of cell death is not known (Hard et al., 1993; Swenberg, 1993). The degenerate cells and cell fragments are exfoliated from the proximal tubule into the tubule lumen forming granular casts (Hard et al., 1993; Swenberg, 1993). The necrosis is accompanied by a dose-specific increase in epithelial cell proliferation in the proximal tubule (Short et al., 1989; Swenberg, 1993). The sustained increase in cell proliferation is capable of promoting spontaneously initiated or chemically initiated cells of the proximal tubule to form preneoplastic and neoplastic lesions (Dietrich and Swenberg, 1991a; Read et al., 1988). The rapid rate of cell proliferation reduces the time available to repair DNA damage, increasing the likelihood of fixing a mutation and promoting clonal expansion of cells containing such mutations, thereby increasing the probability of neoplasia (Loeb, 1989).

1.5. SCOPE AND NATURE OF THIS WORK

The complex tissue, hormonal, and developmental expression patterns observed for members of the rat alpha-2u-globulins makes this gene family an interesting system for studying the evolution of differential gene expression. As a result of this interest, numerous studies have been carried out in attempts to identify and investigate the regulatory sequences and transcription factors that control alpha-2u-globulin expression (see 1.4). Comparative studies require the isolation of genomic clones of a number of differentially expressed genes, a task made difficult by the high degree of similarity between the different alpha-2u-globulin genes. To fully investigate

the genetic basis of the differential hormonal, developmental, and tissue specific expression the members of this gene family a systematic study of the gene family is required.

The genomic organization of the alpha-2u-globulin genes has not been studied in any detail. Somatic cell hybrid panel screening, radioisotopic *in situ* hybridization, and linkage analysis have indicated that the alpha-2u-globulin genes are clustered on rat chromosome 5 (Kurtz, 1981; Nikaido et al., 1982). Although, the alpha-2u-globulins are apparently clustered, lambda genomic clones containing more than one gene have not been isolated, suggesting the repeat unit is potentially larger than the approximately 15 kb commonly accommodated by bacteriophage lambda cloning systems. The work presented here utilizes an integrated approach of molecular biology and cytogenetics to investigate the the organization of the alpha-2u-globulin multigene family.

As a first step toward understanding this multigene family the chromosomal location of the alpha-2u-globulin genes was determined using FISH. The results of this analysis indicated that the alpha-2u-globulin genes were clustered at a single location, permitting an estimate of the size of the locus to be made using pulsed field gel electrophoresis.

The availability of a rat genomic P1 library allowed for the the isolation of three large, non-overlapping genomic clones from the locus. These clones permitted, for the first time, an assessment of the linkage relationships between genes, using both molecular techniques and recently developed high resolution Fiber FISH methods, that would have otherwise been difficult given the sequence similarity of the alpha-2u-globulin genes. Application of the Fiber FISH method was also extended to an analysis of genomic DNA to assess, on a larger scale, the organization of the entire locus.

The alpha-2u-globulin gene cluster was further characterized using differential hybridization methods and diagnostic restriction endonuclease digestion to identify classes, or types, of alpha-2u-globulin genes, and to relate these classes to the overall genomic organization of the locus.

There is evidence from other gene families, such as the globins and the homeotic genes, that chromosomal arrangement has functional significance. The elucidation, therefore, of the organization of the alpha-2u-globulin gene cluster is a requisite step towards the complete understanding of the complex tissue, hormonal, and developmental regulation of the alpha-2u-globulin genes.

CHAPTER 2.
MATERIALS AND METHODS

2.1. TISSUE CULTURE METHODS

2.1.1. CELL LINES

Two cell lines were used in this study. The Rat 2 cell line is derived from a sub-clone of a 5'-bromo-deoxyuridine resistant strain of the Fischer rat fibroblast 3T3-like cell line Rat 1 which was established from embryonic tissue (Topp, 1981). Rat 2 cells lack appreciable levels of thymidine kinase, are highly transfectable and are phenotypically normal. I also used the mouse L-M (TK$^-$) cell line, which is a thymidine kinase mutant, 5'-bromo-deoxyuridine resistant cell line established from the mouse L-M fibroblast cell line (Kit et al., 1963). Both lines were acquired from ATCC.

2.1.2. GROWTH CONDITIONS

Rat 2 and mouse L fibroblasts were routinely cultured in 10 ml of Dulbecco's Modified Eagle Medium (DMEM, BRL)/10% calf serum in 100 mm culture dishes (Nunc). Growth conditions were maintained at 37°C and 5% CO_2 in a humidified tissue culture incubator (Forma). Cells were fed fresh medium every 2 days, and routinely sub-cultured approximately every 8-10 days. To sub-culture, medium was removed from the cells, cells were then washed with 10 ml of phoshate buffered saline (PBS; 2.68 mM KCl, 1.47 mM KH_2PO_4, 137 mM NaCl, 8.06 mM $Na_2HPO_4 7H_2O$). Cells were then detached from the growth surface by treating them with 1 ml of trypsin-EDTA (BRL) for several minutes. Trypsin was inactivated by the addition of 9 ml of DMEM/10% calf serum. The detached cells were seeded in a new dish typically at a 10 fold dilution.

2.2. MOLECULAR BIOLOGICAL METHODS

2.2.1. RECOMBINANT DNA TECHNIQUES

2.2.1.1. BACTERIAL STRAINS

The bacterial strains used in this project, and their relevant genotypes, are listed in Table 2.1.

Table 2.1. Bacterial Strains.

Strain	Relevant Genotype	Reference
DH5 α	*supE*44 Δ*lac*U169 (Φ80 *lacZ*ΔM15) *hsdR*17 *rec*A1 *end*A1 *gyr*A96 *thi*-1 *rel*A1	(1)
DH10B	F⁻ *mcrA* Δ(*mrr-hsdRMS-mcr BC*) (Φ80 *lacZ*ΔM15) Δ*lacX*74 *endA*1 *recA*1 Δ(*ara-leu*)7697 *araD*139 *galUK nupG rpsL*	(2)
NS3529	*recA*1 *mcrA* Δ(*mrr-hsdRMS-mcrBC*) λ*imm*λLP1 λ*imm*434 *nin*5X1-cre	(3)
PLK-17	e14⁻ *mcrA mcrB*1 *lac hsdR*2 *supE*44 *galK*2 *galT*22 *metB*1	(4)

References (1) Woodcock et al., 1989; (2) Grant et al., 1990; (3) Pierce et al., 1992; (4) Kretz et al., 1989.

2.2.1.2. ALPHA-2U-GLOBULIN CONTAINING PLASMIDS

The plasmid 91R91 (Figure 2.1), containing a 5.3 kb Hind III-Bam HI fragment from genomic clone λ91 (Kurtz, 1981) cloned into pBR322 was used as a source for alpha-2u-globulin probes. Plasmid 91R91 contains the entire coding region of alpha-2u-globulin gene 91, approximately 750 bp of upstream sequences, and 400 bp of 3' flanking sequences.

The plasmid pSK2-3 (Figure 2.1) was constructed by subcloning a 3.2 kb Hind III fragment from clone pL 1 (Wang, 1996; Figure 2.2). This construct contains sequences between +750 and +3950 bp of alpha-2u-globulin gene 91, and was used as a probe in genomic fiber FISH experiments.

2.2.1.3. PLASMID DNA PREPARATION

Plasmid DNAs were purified using the alkaline lysis method of Birnboim and Doly with minor modifications (1979). Briefly, bacterial cells grown in 500 ml of Luria Broth (LB; 10 g/l tryptone, 5 g/l yeast extract, 5 g/l NaCl) (Miller, 1972) containing the appropriate antibiotic were pelleted in a GSA rotor (Sorvall) at 5000 rpm at 4°C. Each pellet (from 250 ml of culture) was resuspended in 6 ml of GET (50 mM glucose, 25 mM Tris-HCl (pH 8.0), 10 mM EDTA (pH 8.0)) and 2 ml of lysozyme (20 mg/ml in GET) was added. The samples were incubated on ice for 5 minutes, then 16 ml of 1% SDS in 0.2 M NaOH was added, the contents mixed by gently inversion, and the incubation on ice was continued for another 10 minutes. Following the incubation, 12 ml of ice cold solution III (for 100 ml: 60 ml 5 M potassium acetate, 11.5 ml glacial acetic acid, 28.5 ml H_2O) was added, the contents mixed by inversion and incubated on ice for 10 minutes. The lysate was centrifuged for 30 minutes at 15 000 rpm in a SS34 rotor (Sorvall) at 4°C. The supernatant was removed, and plasmid DNA precipitated by the addition of 0.6 volumes of room temperature isopropanol. DNA was recovered by centrifugation for 30 minutes at 10 000 rpm in a SS34 rotor at room temperature.

The pellet containing plasmid DNA was washed twice with ice cold 70% ethanol, and dried. The pellets from 500 ml of culture were combined and resuspended in 4 ml of TE (10 mM Tris-HCl, 1 mM EDTA (pH 8.0)), to which 4 grams of CsCl and 400 µl of ethidium bromide (10 mg/ml) was added. The solution was mixed and left at room temperature for 15 minutes to allow particulate matter to settle, before centrifuging at 8000 rpm for 10 minutes in a SS34 rotor at room temperature.

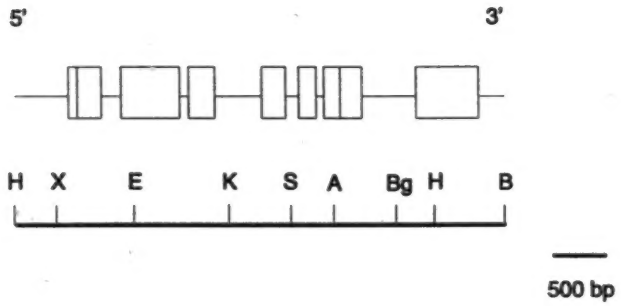
Figure 2.1. Alpha-2u-globulin containing plasmids.

(A) The upper drawing shows the genomic structure of alpha-2u-globulin gene 91. Exons are shown as boxes with the translated regions shaded grey. The second line shows a restriction map of gene 91 indicating restriction sites utilized in this work.

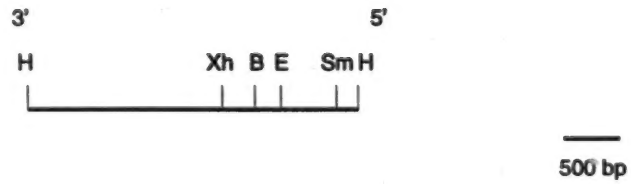
(B) Restriction map of the genomic insert in pSK2-3.

For each panel the restriction sites are: A, Ava I; B, Bam HI; Bg, Bgl II; E, Eco RI; H, Hind III; K, Kpn I; S, Sal I; Sm, Sma I; X, Xba I; Xh, Xho I. Only relevant restriction sites are indicated.

A



B



The supernatant was centrifuged in an ultracentrifuge tube (Beckman, polyallomer) for at least 20 hours at 48 000 rpm in a VTi 65 rotor (Beckman) at 20°C. Following centrifugation, a band corresponding to covalently closed circular plasmid DNA was visualized with long wave ultra-violet light, and recovered by puncturing the tube with an 18 gauge needle attached to a 1 ml syringe and withdrawing the band. Ethidium bromide was extracted with n-butyl alcohol saturated with 4 M NaCl, 10 mM Tris-HCl and 1 mM EDTA (pH 8.0). After dialysis for 16 hours against two changes of dialysis buffer (4 mM NaCl, 1 mM EDTA, 10 mM Tris-HCl (pH 8.0)), the purified DNA was stored at 4°C.

On occasion plasmids were isolated and purified using ion exchange columns (Qiagen) following the protocol supplied by the manufacturer.

2.2.1.4. RESTRICTION ENDONUCLEASE DIGESTIONS

Restriction enzymes were obtained from several commercial suppliers (BRL, NEB, Pharmacia, Boehringer). Restriction enzyme digestions were performed as recommended by the vendor, in the buffers supplied with each enzyme. The exception to this was the partial digestions performed on the P1 clones (see 2.2.4.4) for which universal restriction buffer (33 mM Tris-Acetate (pH 7.9), 66 mM potassium acetate, 10 mM magnesium acetate, 100µg/ml bovine serum albumin, 0.5 mM dithiolthreitol, 4 mM spermidine; Tartoff and Hobbs, 1988) was used.

2.2.1.5. AGAROSE GEL ELECTROPHORESIS

All agarose gel electrophoresis was carried out in gels of the appropriate concentration to resolve the fragments of interest using 0.5x TBE (1x TBE; 90 mM Tris-borate, 2 mM EDTA (pH 8.0)) as the running buffer. Ethidium bromide, at a final concentration 0.6 µg/ml, was typically included in the gel.

2.2.1.6. SOUTHERN BLOTTING

After electrophoresis and photodocumentation, gels containing digested DNA were treated sequentially with depurination solution (0.25 M HCl), and denaturation solution (0.4 M NaOH) for 15 minutes each at room temperature with constant shaking. If the size of the fragments to be blotted was less than 10 kb, the depurination step was ommitted. Capillary transfers to nylon membranes (Hybond-N$^+$, Amersham) were set up according to Southern (1975), using 0.4 M NaOH as the transfer solution. Following transfer, membranes were briefly rinsed in 2x SSPE (20x SSPE; 3.6 M NaCl, 0.2 M sodium phosphate, 0.02 M EDTA pH 8.0), wrapped in plastic wrap and stored at 4°C until required.

2.2.1.7. SLOT BLOTTING

Slot blots were prepared as described by Ausubel et al. (1995). Briefly, DNA samples were prepared in a 100 µl volume containing 10 mM EDTA and 0.4 M NaOH boiled for 5 min. The denatured DNA samples were applied to a damp nylon membrane (Hybond N^+) using a manifold assembly (Tyler). Slots in which samples had been applied were rinsed with 200 µl of 400 mM NaOH before the manifold was disassembled and the membrane rinsed in 2x SSPE. The slot blot was hybridized as discussed below.

2.2.1.8. MEMBRANE PRE-HYBRIDIZATION

Membranes containing immobilized DNA were placed in glass hybridization bottles with 10 ml of hybridization solution (5x SSPE, 5x Denhardt's (100x Denhardt's; 2% bovine serum albumin, 2% Ficoll, 2% polyvinylpyrrolidone), 0.5% SDS (sodium dodecyl sulfate), 20 µg/ml sheared denatured salmon sperm DNA) and prehybridized at 65°C for 1 hour in a hybridization oven (Robbins).

2.2.1.9. PREPARATION OF RADIO-LABELED PROBE DNA

DNA fragments desired for probes in Southern hybridization were generated by digesting the desired template DNA with the appropriate restriction enzyme(s). Restriction fragments were separated in agarose gels using 1x TAE buffer (40 mM Tris-acetate, 1 mM EDTA (pH 8.0)) as the running buffer. The desired fragment was excised from the gel and melted in 3 volumes of 6 M NaI at 50°C. After the agarose slice was completely melted, 20 µl of silica (Sigma) suspension (100 mg/ml in 3 M NaI) was added and the mixture was incubated on ice for 5 minutes. The silica was pelleted by centrifugation for 5 seconds at 14 000 rpm in a microcentrifuge, the supernatant was removed and the pellet resuspended in 500 µl of wash buffer (50 mM NaCl, 10 mM Tris-HCl pH 7.5, 2.5 mM EDTA, 50% v/v ethanol). The silica was pelleted as before, washed again as described, and pelleted once more by centrifugation for 5 seconds at 14 000 rpm in a microcentrifuge. The supernatant was removed from the pellet, and the pellet was resuspended in 20 µl of H_2O and incubated for 5 minutes at 50°C. After the incubation the silica was pelleted by centrifugation for 2 minutes at 14 000 rpm in a microcentrifuge. The supernatant, containing the desired DNA, was transfered to a new microcentrifuge tube. Typically, 8 µl of the recovered DNA solution was used in a random priming labeling reaction with α^{32}P-dCTP (3000 Ci/mmole; 10 µCi/µl, Amersham) using the Quick Prime labeling kit (Pharmacia) according to the manufactures directions.

2.2.1.10. HYBRIDIZATION AND WASH PROTOCOLS

Before use, the probe was diluted with H_2O, so that the final radioactivity added to the 10 ml hybridization solution was between 10^5-10^6 cpm/ml, and denatured by boiling in a water bath for 5 minutes. Hybridization was carried out overnight at 65°C.

Membranes were routinely washed once for 30 minutes in 4x SSPE/0.1% SDS at 65°C, twice for 30 minutes each in 2x SSPE/0.1% SDS at 65°C, and twice for 30 minutes each in 1x SSPE/0.1% SDS at 65°C. On occasion, as signal dictated, additional 15 minute washes with 0.1x SSPE/0.1% SDS at 65°C were carried out to reduce background to acceptable levels. Membranes were then blotted to remove excess wash solution, sealed in plastic bags, and exposed to X-ray film at -80°C.

2.2.1.11. ELECTROPORATION

Single colonies from strains to be used in electroporation were innoculated into 10 ml LB and grown overnight at 37°C with shaking. The 10 ml overnight cultures were used to prime growth in 1 liter of LB. Cultures were grown at 37°C with shaking until the OD_{600} had reached 0.5-1.0 at which time the culture was split into 4 centrifuge bottles each containing 250 ml and centrifuged at 5000 rpm for 15 minutes at 4°C in a GSA rotor. Following centrifugation the pellets were resuspended in 1 volume (250 ml) of sterile cold water and centrifuged as above. The cell pellets were washed twice more using 0.5 volumes (125 ml) of sterile cold water and centrifuged as before. At this point the 4 cell pellets were combined in 0.02 original culture volumes (20 ml) of sterile cold water and centrifuged at 5000 rpm for 15 minutes at 4°C in a SS34 rotor. The final cell pellet was resuspended in 0.002 volumes (2 ml) of filter sterilized cold 10% glycerol and dispensed in 40 µl aliquots, and stored at -70°C.

Electroporation was performed using an Electro-cell manipulator 600 (BTX) in high voltage mode. Capacitance was set to 0, resistance was 129 Ohms (R5), and the charging voltage was 2.45 kV. An aliquot containing 1 ng of plasmid DNA in water or TE was mixed with 40 µl of electro-competent *Escherichia coli* cells on ice, and transfered to a chilled electro-poration cuvette with a 2 mm gap size (BTX). The cuvette was pulsed in the electroporator, and 960 µl of LB was immediately added to the cuvette. The cell mixture was transfered to a sterile tube, and incubated at 37°C for 1 hour with gentle shaking. Appropriate dilutions of the cell mixtures were plated to the desired selective media.

2.2.1.12. IMAGE ANALYSIS PROTOCOL

Photographs of agarose gels, and autoradiographs included in the thesis were electronically scanned using a model GS-670 imaging densitometer (BioRad) and the Molecular Analyist program (BioRad). The digitized images were prepared for presentation using CorelDraw v3.0 (Corel).

2.2.2. ISOLATION OF GENOMIC DNA

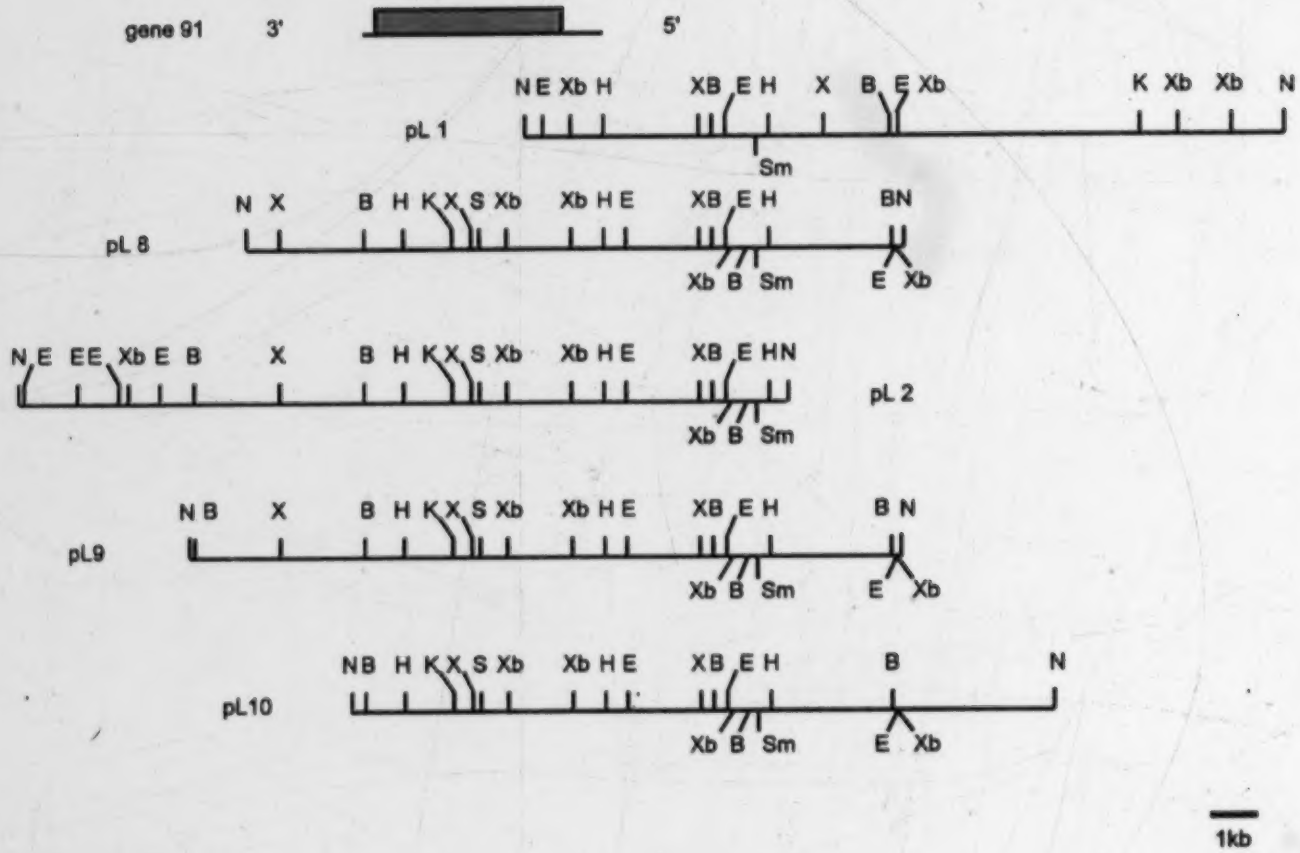
Genomic DNA was isolated from cultured cell lines esentially according to the method outlined by Ausubel et al. (1995) with minor modifications. Briefly, tissue culture cells in 100 mm culture dishes, were washed *in situ* with 8 ml of cold PBS and then treated with 1 ml of digestion buffer (100 mM NaCl, 10 mM Tris-HCl pH 8.0, 25 mM EDTA, 0.5% SDS, and 0.1 mg/ml proteinase K). The cell lysate was scraped into a 10 ml plastic centrifuge tube with the aid of a rubber policeman, and incubated overnight at 50°C. Following incubation, the lysate was extracted with an equal volume of phenol/$CHCl_3$ (1:1) and centrifuged at 8000 rpm for 10 minutes at room temperature in an SS34 rotor. Genomic DNA was precipitated from the supernatant following the addition of 0.5 volume of 7.5 M ammonium acetate and 2 volumes of ice cold 95% ethanol. The DNA was recovered by spooling the DNA onto the hooked end of a glass capillary tube, and washed by twice dipping the DNA into ice cold 70% ethanol. Residual ethanol was allowed to evaporate before the DNA was resuspended in TE. Heating the DNA at 65°C aided in its dissoluton in TE.

2.2.3. RAT LAMBDA GENOMIC LIBRARY ANALYSIS

2.2.3.1. ISOLATION OF LAMBDA GENOMIC CLONES CONTAINING ALPHA-2U-GLOBULIN GENE 91

A rat genomic library in λ DASH II (Stratagene) was screened (Wang, 1996) with an end labelled oligonucleotide which overlaps the transcription initiation site (-8 bp-+32 bp) of alpha-2u-globulin gene 91, in the plasmid 91R91. Single plaque suspensions prepared from positive plaques in the primary screen were rescreened using the PCR with primers derived from the region between -230 bp and +60 of gene 91 (Wang, 1996). This region appears to be somewhat divergent in 13 family members examined (Addison, unpublished observation). A collection of 5 clones (Figure 2.2) were obtained, purified, and subcloned into pBluescript (Wang, 1996). Four of the clones, pL2, pL8, pL9, and pL10 contain the entire alpha-2u-globulin coding region and various lengths of upstream and downstream sequences, and appear, based on restriction endonuclease sites, and partial sequence analysis, to be independent isolates of the same member of the alpha-2u-globulin gene family (Wang, 1996).

Figure 2.2. Restriction Maps of 5 subclones from the rat genomic lambda library. The genomic inserts were subcloned on Not I fragments in to pBluescript. The Hind III-Bam HI fragment from 91R91 which contains alpha-2u-globulin gene 91, is presented at the top of the figure for orientation. Restriction sites shown are: B, Bam HI; E, Eco RI; H, Hind III; K, Kpn; N, Not I; S, Sal I; Sm, Sma I; X, Xho I; Xb, Xba I. The right end of pL 1 is shown in more detail in Figure 3.3.


gene 91
3'
5'
pL 1
N E Xb H
XB E H
X
B E Xb
K Xb Xb N
Sm
pL 8
N X
B H K X S Xb
Xb H E
XB E H
BN
Xb B Sm
E Xb
N E EE Xb E B
X
B H K X S Xb
Xb H E
XB E HN
pL 2
Xb B Sm
pL9
N B
X
B H K X S Xb
Xb H E
XB E H
B N
Xb B Sm
E Xb
pL10
N B H K X S Xb
Xb H E
XB E H
B
N
Xb B Sm
E Xb
1kb

The fifth clone, pL1, only contains a portion of the 5' end of the alpha-2u-globulin gene, and appears to be identical to gene 91 (Wang, 1996).

2.2.3.2. DIFFERENTIAL STRINGENCY SCREENING

Plating cells were prepared by innoculating 50 ml of LB supplemented with 0.2% maltose and 10 mM $MgSO_4$ with a single colony of *Escherichia coli* strain PLK-17. Cells were grown overnight at 30°C with shaking. The culture was centrifuged for 10 min at 2000 rpm in an SS34 rotor to pellet the cells. Following centrifugation the media was decanted, and the cell pellet was gently resuspended in 10 mM $MgSO_4$ and the concentration of cells adjusted to an $OD_{600} = 0.5$. This cell suspension was stored at 4°C until required.

The λ Dash II rat genomic library was diluted to the appropriate titre in SM (5.8 g/l NaCl, 2.0 g/l $MgSO_4 \cdot H_2O$, 50 ml/l 1 M Tris-HCl (pH 7.5), 5.0 ml/l 2% gelatin), mixed with plating cells, and plated in NZY top agar (22 g/l NZCYM broth (Becton Dickinson), 0.7% agarose) on NZY plates (22 g/l NZCYM broth, 15 g/l agar). The plates were allowed to solidify and then incubated for 9 hours at 37°C, and then stored at 4°C.

Duplicate plaque lifts were made from each of six plates, containing approximately 50 000 phage per plate, using nylon membranes (MAGNA, MSI) allowing 2 minutes for the first transfer and 5 minutes for the second lift. The filters were denatured by submersion in 1.5 M NaCl/0.5 M NaOH for 2 minutes, and then neutralized by submersion in 1.5 M NaCl/0.5 M Tris-HCl (pH 8.0) for 5 minutes. Filters were then rinsed briefly in 2x SSC, air dried, and baked at 80°C for 20-30 minutes.

The filters were prehybridized in 250 mM Na_2HPO_4 (pH 7.2), 1 mM EDTA, 7% SDS at 50°C for 1 hour with shaking. Hybridization was performed with 10^5-10^6 cpm/ml of radioactively labelled probe prepared from the 4 kb Hind III fragment of plasmid 91R91 (Figure 2.1), which contains an alpha-2u-globulin gene, overnight at 50°C. Hybridization at 50°C theoretically allows for hybridization to sequences with 30% mismatch to the probe.

Following hybridization, all 12 filters were washed twice for 15 minutes each in 250 mM Na_2HPO_4 (pH 7.2), 1 mM EDTA, 1% SDS at room temperature with shaking. Following the first set of washes the duplicate membranes were differentially washed: one set of 6 filters was washed at low stringency- three washes at 42°C in 4x SSC/0.1% SDS for 15 minutes each with shaking. The remaining set of six filters were washed three times in 0.1x SSC/0.1% SDS for 15 minutes each at 65°C. The high stringency washes only permit the retention of hybrids with 5% or less mismatch to the probe. Following the washes the filters were blotted dry, wrapped in plastic wrap, and exposed to X-ray film at -80°C.

Both filter sets were stripped by boiling in 0.1x SSC/0.1% SDS and reprobed with probes derived from lambda subclone pL1 (Figure 2.2). Hybridization and washes were carried out as described previously, except that the most stringent wash was performed in 2x SSC/0.1% SDS at 65°C. In addition, the filters were hybridized separately with the 500 bp Bgl II-Hind III fragment and the 600 bp Ava I-Bgl II fragment, both derived from adjacent regions in the 3' end of gene 91 (Figure 2.1) to examine divergence in the 3' ends of alpha-2u-globulin genes.

2.2.4. RAT BACTERIOPHAGE P1 LIBRARY ANALYSIS

2.2.4.1. DEVELOPMENT OF PCR PRIMERS AND CONDITIONS

The sequence of exons 2 and 3 from alpha-2u-globulin clone 207 (Dolan et al., 1983) was entered into the PCRPLAN program in the PC/Gene package (Intelligenetics), and a compatible primer pair meeting the criteria in Table 2.2 was obtained.

A primer pair, designated DAMF/DAMR (see Figure 2.3), met the set criteria and was optimized for amplification, using the PCR, of the expected 318 bp product by varying the concentration of $MgCl_2$ in the amplification reaction. PCR reactions were carried out in 30 µl reaction volumes containing: 50 mM KCl, 10 mM Tris-HCl (pH 8.4), 0.2 mM each dNTP, 25 pmol each primer, and 2.5 units Taq polymerase. The $MgCl_2$ concentration was varied from 0-5 mM in 0.5 mM increments. The cycling parameters were: 94°C for 1 minute, 49°C for 45 seconds, and 72°C for 1 minute for 30 cycles performed in a Robocycler 40 (Stratagene).

2.2.4.2. P1 LIBRARY SCREENING

Library screening was performed by Genome Systems Inc. The library is composed of 209 primary pools, each containing on average 550 clones (range 230-1200 clones/pool). To screen the library, all 209 pools were screened by PCR with the oligonucleotide pair DAMF/DAMR with the cycling parameters stated above (2.2.4.1). Positive pools were identified by the presence of a 318 bp PCR product observed following agarose gel electrophoresis. Given the repetitive nature of the alpha-2u-globulin gene, a significant number of the 209 primary pools tested positive for alpha-2u-globulin-containing clones. Three of the positive pools were plated and screened by colony hybridization using the PCR product obtained with the primer pair DAMF/DAMR on a rat genomic template as a probe. One positive clone, of many, from each of the three selected pools was obtained from Genome Systems Inc. as part of the routine screening. Additional positive clones were not purchased.

Table 2.2. PCR Primer Parameters

Parameter	Set Value
Optimal Tm	60°C
Acceptable Tm range	45-75°C
Salt concentration	50 mM
Primer concentration	250 pM
Optimal primer length	22 bp
Acceptable primer length	17-27 bp
Maximum single base repitition	3
Number of bases in 3' GC clamp	2
Acceptable range for %GC	45-60%
Maximum number of bases for primer self-complementarity	3
Maximum stem size in stem-loop formation	4
Maximum % of bases complementary to other template regions	70%
Maximum number of bases complementary between primers	3

Figure 2.3. Exonic sequences of 12 alpha-2u-globulin genes reported in Genbank. Accession numbers are provided to the left of each sequence. The first nucleotide of each exon is marked by arrow heads. Only the coding region of exon 1 is reported. Sequences referred to in the text are underlined and labeled. Nucleotide differences are noted in bold face type. The origin of each sequence is as indicated below.

M26837	liver cDNA clone
M26838	liver cDNA clone
M26835	liver cDNA clone
U31287	liver cDNA clone
M26836	liver cDNA clone
J00737	liver cDNA clone
M12155	liver cDNA clone
J00736	genomic clone from liver derived library
J00738	submaxillary gland cDNA clone
X14552	salivary gland cDNA clone
X14434	genomic clone from liver derived library
X14435	genomic clone from liver derived library

> Exon 1

```
M26837  ATGAAGCTGTTGCTGCTGCTGTGTCTGGGCCTGACACTGGTC
M26838  ATGAAGCTGTTGCTGCTGCTGTGTCTGGGCCTGACACTGGTC
M26835  ATGAAGCTGTTGCTGCTGCTGTGTCTGGGCCTGACACTGGTC
U31287  ATGAAGCTGTTGCTGCTGCTGTGTCTGGGCCTGACACTGGTC
M26836  ATGAAGCTGTTGCTGCTGCTGTGTCTGGGCCTGACACTGGTC
J00737  ------------CTGCTGCTGTGTCTGGGCCTGACACTGGTC
M12155  ----------------------------------------
J00736  ----------------------------------------
J00738  ATGAAGCTGTTGCTGCTGCTGTGTCTGGGCCTGACCCTGGTC
X14552  ATGAAGCTGTTGCTGCTGCTGTGTCTGGGCCTGACCCTGGTC
X14434  ATGAAGCTGTTGCTGCTGCTGTGTCTGGGCCTGACCCTGGTC
X14435  ATGAAGCTGTTGCTGCTGCTGAGTCTGGGCCTGACCCTGGTC

M26837  TGTGGCCATGCAGAAGAAGCTAGTTCCACAAGAGGGAACCTC
M26838  TGTGGCCATGCAGAAGAAGCTAGTTCCACAAGAGGGAACCTC
M26835  TGTGGCCATGCAGAAGAAGCTAGTTCCACAAGAGGGAACCTC
U31287  TGTGGCCATGCAGAAGAAGCTAGTTCCACAAGAGGGAACCTC
M26836  TGTGGCCATGCAGAAGAAGCTAGTTCCACAAGAGGGAACCTC
J00737  TGTGGCCATGCAGAAGAAGCTAGTTCCACAAGCGGGAACCTC
M12155  ----------------------AGTTCCACAAGAGGGAACCTC
J00736  ----------------------------------------
J00738  TGTGGCCATGCAGAAGAAGCTAGTTTCGAGAGAGGGAACCTC
X14552  TGTGGCCATGCAGAAGAAGCTAGTTTCGAGAGAGGGAACCTC
X14434  TGTGGCCATGCAGAAGAGGCTAATTCCGAGAGAGGGAACCTC
X14435  TGTGGCCATGCAGAAGACGCTAGTTCCACAAGAGGGAACCTC
```

Exon 2 >

```
M26837  GATGTGGCTAAGCTCAATGGGGATTGGTTTTCTATTGTCGTG
M26838  GATGTGGACAAGCTCAATGGGGATTGGTTTTCTATTGTCGTG
M26835  GATGTGGCTAAGCTCAATGGGGATTGGTTTTCTATTGTCGTG
U31287  GATGTGGCTAAGCTCAATGGGGATTGGTTTTCTATTGTCGTG
M26836  GATGTGGACAAGCTCAATGGGGATTGGTTTTCTATTGTCGTG
J00737  GATGTGGCTAAGCTCAATGGGGATTGGTTTTCTATTGTCGTG
M12155  GATGTGGCTAAGCTCAATGGGGATTGGTTTTCTATTGTCGTG
J00736  ----------AGCTCAATGGGGATTGGTTTTCTATTGTCGTG
J00738  GATGTGGACAAGCTCAATGGGGATTGGTTTTCTATTGTCGTG
X14552  GATGTGGACAAGCTCAATGGGGATTGGTTTTCTATTGTCGTG
X14434  GATGTGGACAAGCTCAATGGGGATTGGTTTTCTATTGTCGTG
X14435  GATGTGGACAAGCTCAATGGGGATTGGTTTTCTATTGTCGTG

M26837  GCCTCTAACAAAAGAGAAAAGATAGAAGAGAATGGCAGCATG
M26838  GCCTCTGACAAAAGAGAAAAGATAGAAGAGAATGGCAGCATG
M26835  GCCTCTAACAAAAGAGAAAAGATAGAAGAGAATGGCAGCATG
U31287  GCCTCTAACAAAAGAGAAAAGATAGAAGAGAATGGCAGCATG
M26836  GCCTCTGACAAAAGAGAAAAGATAGAAGAGAATGCCAGCATG
J00737  GCCTCTAACAAAAGAGAAAAGATAGAAGAGAATGGCAGCATG
M12155  GCCTCTAACAAAAGAGAAAAGATAGAAGAGAATGGCAGCATG
J00736  GCCTCTAACAAACGAGAAAAGATTGAAGAGAATGGCAGCATG
```

```
J00738  GCCTCTGATAAAAGAGAAAAGATAGAAGAGAACGGCAGCATG
X14552  GCCTCTGATAAAAGAGAAAAGATAGAAGAGAACGGCAGCATG
X14434  GCCTCTAACAAACGAGAAAAGATTGAAGAGAATGGCAGCATG
X14435  GCCTCTAACAAACGAGAAAAGATTGAAGAGAATGGCAGCATG
```

KAW-1

```
M26837  AGAGTTTTTATGCAGCACATCGAIGICIIGGAGAAIICCTTA
M26838  AGAGTTTTTATGCAGCACATCGATGTCTTGGAGAATICCIIA
M26835  AGAGTTTTTATGCAGCACATCGATGTCTTGGAGAATTCCTTA
U31287  AGAGTTTTTATGCAGCACATCGATGTCTTGGAGAATTCCTTA
M26836  AGAGTTTTTATGCAGCACATCGATGTCTTGGAGAATTCCTTA
J00737  AGAGTTTTTATGCAGCACATCGATGTCTTGGAGAATTCCTTA
M12155  AAAGTTTTTATGCAGCACATCGATGTCTTGGAGAATTCCTTA
J00736  AGAGTTTTTATGCAGCACATCGATGTCTTGGAGAATTCCTTA
J00738  AGAGTTTTTGTGCAGCACATCGATGTCTTGGAGAATTCCTTA
X14552  AGAGTTTTTGTGCAGCATATCGATGTCTTGGAGAATTCCTTA
X14434  AGAGTTTTTATGCAGCACATCGATGTCTTGGAGAATTCCTTA
X14435  AGAGTTTTTATGCAGCACATCGATGTCTTGGAGAATTCCTTA
```

DAMF Exon 3 >

```
M26837  GGCTTCAAGTTCCGTATTAAGGAAAATGGAGAGTGCAGGGAA
M26838  GGCIICAAGIICCGTATTAAGGAAAATGGAGAGTGCAGGGAA
M26835  GGCTTCAAGTTCCGTATTAAGGAAAATGGAGAGTGCAGGGAA
U31287  GGCTTCAAGTTCCGTATTAAGGAAAATGGAGAGTGCAGGGAA
M26836  GGCTTCAAGTTCCGTATTAAGGAAAATGGAGAGTGCAGGGAA
J00737  GGCTTCAAGTTCCGTATTAAGGAAAATGGAGAGTGCAGGGAA
M12155  GGCTTCAAGTTCCGTATTAAGGAAAATGGAGAGTGCAGGGAA
J00736  GGCTTCAAGTTCCGTATTAAGTAAAATGGAGAGTGCAGGGAA
J00738  GGCTTCACGTTCCGTATTAAGGAAAATGGAGTGTGCACAGAA
X14552  GGCTTCACGTTCCGTATTAAGGAAAATGGAGTGTGCACAGAA
X14434  GGCTTCAAGTTGTGTATTAAGGAAAATGGAGAGTGCAGGAAA
X14435  GGCTTCAAGTTCCGTATTAAGGAAAATGGAGAGTGCAGGGAA
```

DAMR

```
M26837  CTATATTTGGTTGCCTACAAAACGCCAGAGGAIGGCGAAIAT
M26838  CTATATTTGGTTGCCTACAAAACGCCAGAGGATGGCGAATAT
M26835  CTATATTTGGTTGCCTACAAAACGCCAGAGGATGGCGAATAT
U31287  CTATATTTGGTTGCCTACAAAACGCCAGAGGATGGCGAATAT
M26836  CTATATTTGGTTGCCTACAAAACGCCAGAGGATGGCGAATAT
J00737  CTATATTTGGTTGCCTACAAAACGCCAGAGGATGGCGAATAT
M12155  CTATATTTGGTTGCCTACAAAACGCCAGAGGATGGCGAATAT
J00736  CTATATTTGGTTGCCTACAAAACGCCAGAGGATGGCGAATAT
J00738  TTTTCTTTGGTTGCCGACAAAACAGCAAAGGATGGCGAATAT
X14552  TTTTCTTTGGTTGCCGACAAAACAGCAAAGGATGGCGAATAT
X14434  CTATATTCAGTTGCCTACAAAACGCCAAAGATTGGCGAATAT
X14435  CTATATTTGGTTGCCTACAAAACGCCAGAGGATGGCGAATAT
```

Exon 4 >

```
M26837  TTTGTTGAGTATGACGGAGGGAATACATTTACTATACTTAAG
M26838  TTTGTTGAGTATGACGGAGGGAATACATTTACTATACTTAAG
M26835  TTTGTTGAGTATGACGGAGGGAATACATTTACTATACTTAAG
U31287  TTTGTTGAGTATGACGGAGGGAATACATTTACTATACTTAAG
M26836  TTTGTTGAGTATGACGGAGGGAATACATTTACTATACTTAAG
J00737  TTTGTTGAGTATGACGGAGGGAATACATTTACTATACTTAAG
M12155  TTTGTTGAGTATGACGGAGGGAATACATTTACTATACTTAAG
J00736  TTTGTTGAGTATGACGGAGGGAATACATTTACTATACTGAAG
J00738  TTTGTTGAGTATGACGGAGAAAATACATTTACTATACTGAAG
X14552  TTTGTTGAGTATGACGGAGAGAATACATTTACTATACTTAAG
X14434  TTTCTTGAGTATGACGGAGGGAATACATTTACTATACTTAAG
X14435  TTTGTTGAGTATGACGGAGGGAATACATTTACTATACTGAAG
```

Vsp I

```
M26837  ACAGACTATGACAGATATGTCATGTTTCATCTCAIIAAITTC
M26838  ACAGACTATGACAGATATGTCATGTTTCATCTCATTAATTTC
M26835  ACAGACTATGACAGATATGTCATGTTTCATCTCATTAATTTC
U31287  ACAGACTATGACAGATATGTCATGTTTCATCTCATTAATTTC
M26836  ACAGACTATGACAGATATGTCATGTTTCATCTCATTAATTTC
J00737  ACAGACTATGACAGATATGTCATGTTTCATCTCATTAATTTC
M12155  ACAGACTATGACAGATATGTCATGTTTCATCTCATTAATTTC
J00736  ACAGACTATGACAGATATGTCATGTTTCATCTCATTAATTTC
J00738  ACAGACTATGACAATTATGTCATGTTTCATCTCGTTAATGTC
X14552  ACAGACTATGACAATTATGTCATGTTTCATCTCGTTAATGTC
X14434  ACAGACTATGAAAGATATGTCATGTTTCATCTCGTTAATGTC
X14435  ACAGACTATGACAGATATGTCATGTTTCATCTCATTAATTTC
```

SstI Exon 5 >

```
M26837  AAGAACGGGGAAACCTTCCAGCTGATGGTGCTCTACGGCAGA
M26838  AAGAACGGGGAAACCTTCCAGGCGATGGTGCTCTACGGCAGA
M26835  AAGAACGGGGAAACCTTCCAGCTGATGGTGCTCTACGGCAGA
U31287  AAGAACGGGGAAACCTTCCAGCTGATGGTGCTCTACGGCAGA
M26836  AAGAACGGGGAAACCTTCCAGGCGATGGTGCTCTACGGCAGA
J00737  AAGAACGGGGAAACCTTCCAGCTGATGGTGCTCTACGGCAGA
M12155  AAGAACGGGGAAACCTTCCAGCTGATGGTGCTCTACGGCAGA
J00736  AAGAACGGGGAAACCTTCCAGGCGATGGTGCTCTACGGCAGA
J00738  AACAACGGGGAAACATTCCAGCTGATGGAGCICTACGGCAGA
X14552  AACAACGGGGAAACCTTCCAGCTGATGGAGCTCTATGGCAGA
X14434  AACAATGGGGAAGCCTTCCAGCTGATGGAGCTCTATGGCAGA
X14435  AAGAACGGGGAAACCTTCCAGGCGATGGTGCTCTACGGCAGA

M26837  ACAAAGGATCTGAGTTCAGACATCAAGGAAAAGTTTGCAAAA
M26838  ACAAAGGATCTGAGTTCAGACATCAAGGAAAAGTTTGCAAAA
M26835  ACAAAGGATCTGAGTTCAGACATCAAGGAAAAGTTTGCAAAA
U31287  ACAAAGGATCTGAGTTCAGACATCAAGGAAAAGTTTGCAAAA
M26836  ACAAAGGATCTGAGTTCAGACATCAAGGAAAAGTTTGCAAAA
```

```
J00737  ACAAAGGATCTGAGTTCAGACATCAAGGAAAAGTTTGCAAAA
M12155  ACAAAGGATCTGAGTTCAGACATCAAGGAAAAGTTTGCAAAA
J00736  ACAAAGGATCTGAGTTCAGACATCAAGGAAAAGTTTGCAAAA
J00738  ACAAAGGATCTGAGTTCAGACATCAAGGAAAAGTTTGCAAAA
X14552  ACAAAGGATCTGAGTTCAGACATCAAGGAAAAGTTTGCAAAA
X14434  ACAAAGGATCTGAGTTCAGACATCAAGGAAAAGTTTGCAAAA
X14435  ACAAAGGATCTGAGTTCAGACATCAAGGAAAAGTTTGCAAAA
```

KAW 2

```
M26837  CTATGTGAGGCGCATGGAATCACTAGGGACAATATCATTGAT
M26838  CTATGTGAGGCGCATGGAATCACTAGGGACAATATCATTGAT
M26835  CTATGTGAGGCGCATGGAATCACTAGGGACAATATCATTGAT
U31287  CTATGTGAGGCGCATGGAATCACTAGGGACAATATCATTGAT
M26836  CTATGTGAGGCGCATGGAATCACTAGGGACAATATCATTGAT
J00737  CTATGTGAGGCGCATGGAATCACTAGGGACAATATCATTGAT
M12155  CTATGTGAGGCGCATGGAATCACTAGGGACAATATCATTGAT
J00736  CTATGTGAGGCGCATGGAATCACTAGGGACAATATCATTGAT
J00738  CTATGTGTGGCACATGGAATCACTAGGGACAATATCATTGAC
X14552  CTATGTGTGGCACATGGAATCACTAGGGACAATATCATTGAC
X14434  CTATGTGAGGCGCATGGAATCACTAGGGACAATATCATTGAC
X14435  CTATGTGAGGCGCATGGAATCACTAGGGACAATATCATTGAT
```

Exon 6 >

```
M26837  CTAACCAAGACTGATCGCTGTCTCCAGGCCCGAGGATGAAGA
M26838  CTAACCAAGACTGATCACTGTCTCCAGGCCCGAGGATGAAGA
M26835  CTAACCAAGACTGATCGCTGTCTCCAGGCCCGAGGATGAAGA
U31287  CTAACCAAGACTGATCGCTGTCTCCAGGCCCGAGGATGAAGA
M26836  CTAACCAAGACTGATCACTGTCTCCAGGCCCGAGGATGAAGA
J00737  CTAACCAAGACTGATCGCTGTCTCCAGGCCCGAGGATGAAGA
M12155  CTAACCA-----------------------------------
J00736  CTAACCAAGACTGATCGCTGTCTCCAGGCCCGAGGATGAAGA
J00738  CTAACCAAGACTGATCGCTGTCTCCAGGCCCGAGGTTGAAGA
X14552  CTAACCAAGACTGATCGCTGTCTCCAGGCCTGAGGTTGAAGA
X14434  CTAACCAAGACTGATCGCTGTCTCCAGGCCCGAGGATGAACA
X14435  CTAACCAAGACTGATCGCTGTCTCCAGGCCCGAGGATGAAGA
```

Exon 7 >

```
M26837  AAGGCCTGAGCCTCCAGTGCTGAGTGGAGACTTCTCACCAGG
M26838  AAGGCCTGAGACTCCAGTGCTGAGTGGAGACTTCTCACCAGG
M26835  AAGGCCTGAGCCTCCAGTGCTGAGTGGAGACTTCTCACCAGG
U31287  AAGGCCTGAGCCTCCAGTGCTGAGTGGAGACTTCTCACCAGG
M26836  AAGGCCTGAGACTCCAGTGCTGAGTGGAGACTTCTCACCAGG
J00737  AAGGCCTGAGCCTCCAGTGCTGAGTGGAGACTTCTCACCAGG
M12155  ----------------------------------------
J00736  AAGGCCTGAGCCTCCAGTGCTGAGTGGAGACTTCTCACCAGG
J00738  AAGGCCTGAGCCTCCAGTGCTGAGTGGAGACTTCTCACCAGG
X14552  AAGGCCTGAGCCTCCAGTGCTGAGTGGAGACTTCTCACCAGG
X14434  CAGGACTGAGCCTCCAGTGCTGAGTGGAGACTTCTCACCAGG
```

```
X14435  AAGGCCTGAGCCTCCAGTGCTGAGTGGAGACTTCTCACCAGG

M26837  ACTCTAGCATCACCATTTCCTGTCCATGGACATCCTGAGCAA
M26838  ACTCTAGCATCACCATTTCCTGTCCATGGACATCCTGAGCAA
M26835  ACTCTAGCATCACCATTTCCTGTCCATGGACATCCTGAGCAA
U31287  ACTCTAGCATCACCATTTCCTGTCCATGGACATCCTGAGCAA
M26836  ACTCTAGCATCACCATTTCCTGTCCATGGACATCCTGAGCAA
J00737  ACTCTAGCATCACCATTTCCTGTCCATGGACATCCTGAGCAA
M12155  ----------------------------------------
J00736  ACTCTAGCATCACCATTTCCTGTCCATGGACATCCTGAGCAA
J00738  ACTCCAGCATTACCATTTCCTGTCCATGGACATCCTGAGCAA
X14552  ACTCCAGCATTACCATTTCCTGTCCATGGACATCCTGAGCAA
X14434  ACTCCAGCATCACCATTTCCTGTCCACGGACATCCTGAGCAA
X14435  ACTCCAGCATCACCATTTCCTGTCCATGGACATCCTGAGCAA

M26837  ATTCTGCGATCTGATTTCCATCCTCTGTCAAGAAAAGTGCAA
M26838  ATTCTGCGATCTGATTTCCATCCTCTGTCAAGAAAAGTGCAA
M26835  ATTCTGCGATCTGATTTCCATCCTCTGTCAAGAAAAGTGCAA
U31287  ATTCTGCGATCTGATTTCCATCCTCTGTCAAGAAAAGTGCAA
M26836  ATTCTGCGATCTGATTTCCATCCTCTGTCAAGAAAAGTGCAA
J00737  ATTCTGCGATCTGATTTCCATCCTCTGTCAAGAAAAGTGCAA
M12155  ----------------------------------------
J00736  ATTCTGCGATCTGATTTCCATCCTGTCTCAAGAAAAGTGCAA
J00738  ATTCTGCGATCTGATTTCCATCCTGTCTCAAGAAAAGTGCAA
X14552  ATTCTGCGATCTGATTTCCATCCTGTCTCAAGAAAAGTGCAA
X14434  ATTCTGCGTTCTGATTTCCATCCTGTGTCACGAAAAGTGCAA
X14435  ATTCTGCGATCTGATTTCCATCCTGTCTCACGAAAAGTGCAA

M26837  TCCTGGTCTCTCCAGCATCTTCCCTAG-TTCCCAGGACAACA
M26838  TCCTGGTCTCTCCAGCATCTTCCCTAG-TTCCCAGGACAACA
M26835  TCCTGGTCTCTCCAGCATCTTCCCTAG-TTCCCAGGACAACA
U31287  TCCTGGTCTCTCCAGCATCTTCCCTAGGTTCCCAGGACAACA
M26836  TCCTGGTCTCTCCAGCATCTTCCCTAG-TTCCCAGGACAACA
J00737  TCCTGGTCTCTCCAGCATCTTCCCTAG-TTCCCAGGACAACA
M12155  ----------------------------------------
J00736  TCCCGGTCTCTCCAGCATCTTCCCTAG-TTCCCAGGACAACA
J00738  TCCTGGTCTCTCCAGCATCTTCCCTAG-TTCCCAGGACAACA
X14552  TCCTGGTCTCTCCAGCATCTTCCCTAG-TTCCCAGGACAACA
X14434  TCCTCGTCACTCCAGCATCTTCCCTAG-ATCCCAGGATAACA
X14435  TCCCGGTCTCTCCAGCATCTTCCCTAG-TTCCCAGGACAACA

M26837  CATCGAGAATTAAAAGCTTTCTTAAATTTCCTTGGCCCCCAC
M26838  CATCGAGAATTAAAAGCTTTCTTAAATTTCCTTGGCCCCCAC
M26835  CATCGAGAATTAAAAGCTTTCTTAAATTTCCTTGGCCCCCAC
U31287  CATCGAGAATTAAAAGCTTTCTTAAATTTCCTTGGCCCCCAC
M26836  CATCGAGAATTAAAAGCTTTCTTAAATTTCCTTGGCCCCCAC
J00737  CATCGAGAATTAAAAGCTTTCTTAAATTTCCTTGGCCCCCAC
```

```
M12155  ------------------------------------------------
J00736  CATCGAGAATTAAAAGCTTTCTTAAATTTCCTTTGCCCCCAC
J00738  CATCGAGAATTAAAAGCTTTCTTAAATTTCCTTTGCCCCCAC
X14552  CATCGAGAATTAAAAGCTTTCTTAAATTTCCTTTGCCCCCAC
X14434  CATCGAAAATTAAAAGCTTTCTTAAATTTCCTTTGCCCCCAC
X14435  CAGCGAGAATTAAAAGCTTTCTTAAATTTCCTTTGCCCCCAC

M26837  CCATGATCATTCCGCACAAATATCTTGCTCTGCAGTTCAATA
M26838  CCATGATCATTCCGCACAAATTTCTTGCTCTGCAGTTCAATA
M26835  CCATGATCATTCCGCACAAATATCTTGCTCTGCAGTTCAATA
U31287  CCATGATCATTCCGCACAAATATCTTGCTCTGCAGTTCAATA
M26836  CCATGATCATTCCGCACAAATTTCTTGCTCTGCAGTTCAATA
J00737  CCATGATCATTCCGCACAAATATCTTGCTCTGCAGTTCAATA
M12155  ------------------------------------------------
J00736  CCATGATCATTCCGCACAAATTTCTTGCTCTGCAGTTCAATA
J00738  TCATGATCATTCCGCACAAATTTCTTGCTCTGCAGTGCAATA
X14552  CCATGATCATTCCGCACAAATTTCTTGCTCTGCAGTGCAATA
X14434  CCATGATGATTCTGCACAAATTTCTTGTTCTGCAGTTCAATA
X14435  CCATGATCATTCCGCACAAATTTCTTGCTCTGCAGTTCAATA

M26837  AATGATTACCCTTGCACTT
M26838  AATGATTACCCTTGCACTT
M26835  AATGATTACCCTTGCACTT
U31287  AATGATTACCCTTGCACTT
M26836  AATGATTACCCTTGCACTT
J00737  AATGATTACCCTTGCACTT
M12155  --------------------
J00736  AATGATTACCCTTGCACTT
J00738  AATGATTACCCTTGCACTT
X14552  AATGATTACCCTTGCACTT
X14434  AATGATTACCCTTGCACTT
X14435  AATGATTACCCTTGCACTT
```

2.2.4.3. ISOLATION OF P1 DNA

Strains harboring P1 plasmids were streaked out on LB solid plating medium (LB liquid broth with 1.5% agar) supplemented with kanamycin at a final concentration of 25 μg/ml. Individual colonies were picked and inoculated into 5 ml of LB supplemented with kanamycin at a final concentration of 25 μg/ml and cultured overnight at 37°C with shaking. The 5 ml overnight cultures were used to prime growth in 500 ml of LB with kanamycin at a concentration of 25 μg/ml.

In one isolation protocol utilized, after 1hour of growth, the culture volume was doubled with an equal volume of fresh media, and isopropylthio-β-D-galactoside (IPTG) was added to a final concentration of 1mM to increase the copy number of the P1 plasmid by inducing the lytic replicon. Growth was continued for 5 hours, after which time P1 DNA was isolated as described in 2.2.1.2., but scaled up to accomodate the larger culture volume.

Several other growth strategies were used to increase the yield and quality of P1 DNA. Included among these were growth as above, but in the absence of IPTG, and growth in the rich medium TB (Terrific Broth: 12 g/900 ml tryptone, 24 g/900 ml yeast extract, 4 ml/900 ml glycerol autoclaved, cooled and supplemented with 100 ml of 0.17 M KH_2PO_4/0.72 M K_2HPO_4 after autoclaving) with and without IPTG induction. However, the highest yeilds of P1 DNA were obtained after transfering the P1 plasmids from their original host strain, NS3529, to DH10B by electroporation (2.2.1.11), and isolation from 500 ml of LB using the method described in 2.2.1.2.

2.2.4.4. P1 CLONE MAPPING

The P1 clones were mapped using partial digestions and indirect end-labeling with T7, or SP6 promoter oligonucleotide probes (Ausubel et al., 1995). P1 clones (10μg) were digested in a 100 μl volume of 1x universal buffer with Not I (P1 clones: p2860, and p2862), or Sfi I (P1 clone p2861) to linearize the clone. The 100 μl master digest is sufficient for mapping the sites for 4 restriction enzymes, 20 μl (2 μg) being required for each. An aliquot of 4 μl of master digest was distributed to each of a set of 5 microfuge tubes, held on ice, for each target restriction enzyme (Sal I, Sma I, Sac I, and Pme I). Additional sites for Not I and Nru I were also positioned using double digests. Universal restriction buffer (1X) was added to each of the five microfuge tubes so that the final volume in tube 1 was 15 μl, tubes 2-4 was 10 μl, and tube 5 was 5 μl. A total of 5 Units of the appropriate enzyme was added to tube 1.

After setting up the reactions, 5 µl from tube 1 was transfered to tube 2, mixed, and centrifuged briefly to collect the contents to the bottom of the tube. The sequential transfer of 5 µl to successive tubes continued as described until all 5 tubes contained 10 µl, and were sitting on ice. Restriction digests were initiated by transfering the tubes to a heating block set to a temperature appropriate for each enzyme. Digestion was carried out for 7 minutes, at which time the contents of the 5 tubes were pooled on ice and EDTA (pH 8.0) was added to a final concentration of 100 mM.

The restriction enzyme reactions were mixed with an equal volume of molten 1% low melt agarose, loaded into the wells of a 1% pulsed field gel prepared in 0.5x TBE; allowed to solidify, and electrophoresed in a CHEF DR III pulsed field gel apparatus. The electrophoresis conditions were as follows: 0.5x TBE running buffer at 14°C, 120° re-orientation angle, and a 1-10 second linearly ramped switch time. Electrophoresis was carried out for 24 hours, then the gel was stained, photographed, and transferred by Southern blot.

2.2.5. GENOMIC CLONE TYPING

2.2.5.1. DEVELOPMENT OF PCR PRIMERS AND CONDITIONS

The primer pair KAW1 (5' GATGTCTTGGAGAATTCC 3')/KAW2 (5' CAATGATATTGTCCCTAGTG 3') (Wang, 1996), generated from sequences in the second and fifth exons respectively (see Figure 2.3), were used in the PCR to amplify sequences from genomic DNA, and clones containing genomic segments. PCR was carried out in a volume of 30 µl containing 50 mM KCl, 10 mM Tris-HCL (pH8.4), 0.1 mg/ml gelatin, 1.5 mM $MgCl_2$, 200 µM each dNTP, 25 pmol of each KAW1 and KAW2, and 2.5 Units Taq DNA polymerase. Reactions using cloned segments of genomic DNA as template were performed using 1 ng of linearized DNA (Sfi I digested P1 clones, Not I digested pL series clones, and Bam HI digested 91R91). PCR of isolated genomic DNA was performed using 75 ng of Not I digested DNA isolated from the liver of a male rat.

Digested DNA isolated from a pulsed field gel also served as a template in some reactions. Briefly, agarose blocks containing rat genomic DNA were prepared, digested with Nru I, and fractionated by pulsed field gel electrophoresis as described below (section 2.3). Following electrophoresis, one lane containing Nru I digested DNA was excised from the gel, wrapped in plastic wrap, and stored at 4°C while the remainder of the gel was subject to Southern blot analysis to identify the position of restriction fragments that hybridize with a probe generated from the alpha-2u-globulin gene containing, 4 kb Hind III fragment, of 91R91 (see Figure 2.1). Gel slices (2 mm thick) were cut from the un-blotted gel lane, and DNA was isolated as outlined in

section 2.2.1.9 for the preparation of probe DNA. An aliquot representing 1/3 of the isolated product was used as a template in PCR reactions.

Thermocycling was performed in a Robocycler 40 for a total of 30 cycles. The first cycle of 95°C for 5 minutes, 50°C for 1 minute, 72°C for 2.5 minutes was followed by 29 cycles of 94°C for 1minute, 50°C for 1 minute and 72°C for 2.5 minutes.

2.2.5.2. ANALYSIS OF PCR PRODUCTS

PCR products were analyzed by restriction endonuclease digestion with the enzymes Sst I and Vsp I, which have been used to distinguish different classes of alpha-2u-globulin genes (Gubits et al., 1984; Wang, 1996). PCR products were digested with both Sst I and Vsp I separately, and in combination. The digests were performed in triplicate, and the products resolved on a 2% agarose gel, and subject to Sothern blot hybridization using the uncut PCR product obtained from plasmid 91R91 as a probe.

2.2.5.3. DENSITOMETRIC ANALYSIS OF PCR SOUTHERN BLOT

The resulting autoradiogram was subject to densitometric analysis using a model GS-670 (BioRad) imaging densitometer and the Molecular Analyist program (BioRad) to quantify the the hybridization signal produced by the restriction fragments in each digest. The integrated area under the peaks produced in the densitometric profile for each lane on the autoradiogram were used as a measure of the proportion of each of the classes of alpha-2u-globulin genes.

2.3. PULSED FIELD GEL ANALYSIS OF GENOMIC DNA

2.3.1. PREPARATION OF PULSED FIELD GEL BLOCKS

Agarose blocks containing rat or mouse genomic DNA for pulsed field gel analysis, were prepared according to Birren and Lai (1993). Rat 2 or mouse L fibroblasts cultured in T150 flasks (Corning) were washed with 10 ml of PBS and then trypsinized with 3 ml of trypsin-EDTA to remove the cells from the growth surface. Detached cells were resuspended in 10 ml of DMEM/10% calf serum and centrifuged at 1000 rpm for 5 minutes in a swinging bucket centrifuge (Jouan). The medium was removed and the cells were resuspended at a concentration of 1.3×10^7 cells/ml. The cell suspension was mixed with an equal volume of 1% low melt agarose (Sea Plaque, FMC) in TE, and 200 µl aliquots were dispensed into a block forming mold. The agarose blocks were cooled at 4°C for about 10 minutes until

completely solidified. The blocks were carefully removed from the molds using a sterile spatula, and placed into a 100 mm petri dish containing sufficient ESP (0.5 M EDTA pH 8.0, 1% lauroyl sarcosine, 0.5 mg/ml proteinase K (Boehringer)) to completely cover the blocks. The blocks were incubated at 50°C for 2 days with gentle shaking.

Blocks not required immediately were stored in ESP at 4°C in an air-tight box, after having sealed the edges of the petri dish with parafilm and wrapping the dish with plastic wrap. When required, the blocks were removed from ESP and washed twice for two hours each with TEP (10 mM Tris-HCl pH 8.0, 1 mM EDTA pH 8.0, 1 mM PMSF), and three times for two hours each in TE. Agarose blocks were stored in TE at 4°C in an air-tight box after sealing the edge of the petri dish with parafilm and wrapping the dish with plastic wrap.

2.3.2. RESTRICTION ENDONUCLEASE DIGESTION OF PULSED FIELD AGAROSE BLOCKS

Agarose blocks were taken out of TE and cut into thirds to form plugs and transferred to a 2ml round bottomed microfuge tube. Each plug contained approximately 3 μg of genomic DNA. Plugs were incubated in 250 μl of restriction endonuclease buffer (according to supplier's directions for the specific restriction enzyme) supplemented with bovine serum albumin at a final concentration of 100 μg/ml for 30 minutes on ice. The buffer was exchanged with 250 μl of fresh buffer containing 20 units of the appropriate restriction enzyme. The tubes were held on ice for 30 minutes to allow for diffusion of the enzyme into the plug, without loss of activity, before incubating the tube at 37°C for approximately 16 hours.

2.3.3. ELECTROPHORETIC SEPARATION

Following digestion, the restriction enzyme buffer was removed from the tubes, and the plugs were rinsed with 500 μl of TE. After a brief rinse the plugs were incubated for 30 minutes, on ice, in 500 μl of fresh TE. Plugs were then inserted into the wells of a 1% agarose gel prepared in 0.5x TBE with the aid of a sterile spatula. Gels were electrophoresed using 0.5x TBE as a running buffer in a CHEF DR III pulsed field gel apparatus (Biorad). Specific electrophoresis conditions are provided in the legends of the appropriate figures. Southern blot hybridization was performed on the pulsed field gels as previously described (section 2.2.1.6).

2.4. FLUORESCENT *IN SITU* HYBRIDIZATION

2.4.1. PREPARATION OF METAPHASE SPREADS

Tissue culture medium was removed from a near confluent T 75 flask of Rat 2 fibroblasts, or mouse L cell fibroblasts. Cells were washed with PBS and treated with 3 ml of trypsin-EDTA. The detached cells were resuspended in 10 ml of DMEM/10% calf serum, and 3 ml of this cell suspension was used to seed a T 150 flask containing 27 ml of DMEM/10% calf serum). After approximately 24 hours of growth, the medium was removed and replaced with 30 ml of fresh DMEM/10% calf serum supplemented with colcemid (BRL) at a final concentration of 0.1 μg/ml. The culture was incubated until a significant proportion of the cells had become mitotic, as determined by microscopic observation of cell rounding. Normally, treatment for 4 hours with colcemid was required to attain a high mitotic index.

The medium was removed from colcemid treated cultures and the cells were washed with PBS and treated with trypsin-EDTA. Detached cells were resuspended in 12 ml of DMEM/10% calf serum and transferred to a 15 ml conical centrifuge tube. Cells were pelleted by centrifugation at 800 rpm for 5 minutes in a swinging bucket centrifuge. The supernatant was removed and the cell pellet was resuspended in 15 ml of 0.075M KCl and incubated at 37°C for 10 minutes to swell the cells. The swollen cells were prefixed by the addition of a few drops of carnoy's fixative (methanol:glacial acetic acid, 3:1) and then pelleted by centrifugation at 800 rpm for 10 minutes in a swinging bucket centrifuge. The cell pellet was resuspended in 15 ml of carnoy's fixative and centrifuged at 800 rpm for 10 minutes a total of three times. The final cell pellet was resuspended in a small volume of fixative such that a milky white cell suspension resulted (Usually 1-2 ml of fixative). This suspension was stored at -20°C.

One or two drops of the fixed cell suspension was dropped from a pasteur pipet onto wet, ice-cold slides. Excess fluid was dabbed from the slides with a paper towel, and the slides were placed to steam dry on a test tube rack in a 65°C water bath. Slides were desiccated for 3 days at room temperature before use. Slides could be stored at -70°C, provided they were desiccated again overnight before use.

2.4.2. PREPARATION OF BIOTIN LABELED PROBES

Probes were labelled using the random primer labelling method (Feinberg and Vogelstein, 1983). Plasmid DNA (0.5 μg) in a volume of 16.5 μl was boiled for 5 minutes, and then rapidly cooled on ice. The final volume was

brought to 25 µl by the addition of 2.5 µl of dNTP mix (0.5 mM each dATP, dCTP, dGTP), 2.5 µl of 10x random primer buffer (0.5 M Tris-HCl pH 7.5, 0.1 M $MgCl_2$, 10 mM dithiolthreitol, 0.5 mg/ml bovine serum albumin), 1.25µl of 1 mM biotin-16-dUTP (Boehringer), 1.25 µl random hexamers (150 ng/µl, Pharmacia), and 1µl of Klenow fragment of *E. coli* DNA polymerase I (5 U/µl). The reaction was incubated overnight at 37°C, and stopped by the addition of 1µl of 0.5 M EDTA (pH 8.0) and incubation at 65°C for 10 minutes.

Probes were purified by passing them over a Sephadex G-50 (Pharmacia) column using TE as the elution buffer. Columns were prepared in 1 ml plastic pasteur pipets plugged with glass wool. Before the probes were loaded onto the column, 2 µl of 5% dextran blue (mw 2 x 10^6) was added to follow the progress of the probe through the column. The salt concentration of the eluent containing the probe DNA was adjusted to 300 mM by the addition of 1/10 volume of 3 M sodium acetate (pH 5.2). Probe DNA was precipitated overnight at -20°C after the addition of 2 volumes of ice cold 95% ethanol and centrifugation at 14 000 rpm in a microcentrifuge. The supernatant was carefully removed, the pellet washed with ice cold 70% ethanol, and then dried *in vacuo*. Pellets containing probe DNA were resuspended in hybridization buffer (50% formamide, 2x SSC, 10% dextran sulfate, 40 mM sodium phosphate, 0.1% SDS, 1x Denhardt's, and 100 µg/ml sonicated salmon sperm DNA (pH 7.0)) at a concentration of 0.5 ng/µl. Probes were denatured by heating at 70°C for 5 minutes and rapid cooling on ice before use in *in situ* hybridization.

2.4.3. *IN SITU* HYBRIDIZATION

Slides were treated with RNase A (100 µg/ml in 2x SSC pH 7.0) by pipeting 200 µl on to each slide, covering with a 24 x 50 mm coverslip and incubating at 37°C in a moist chamber prepared from a 150 mm plastic petri dish with a damp paper towel cut to fit the floor. After 1 hour coverslips were removed and the slides were rinsed twice in 2x SSC (pH 7.0) for approximately 30 seconds each, dehydrated in 70% and 95% ethanol for 30 seconds each, and then air-dried. Prior to hybridization, the DNA was denatured in 70% formamide/2x SSC (pH 7.0) at 70°C for 3 minutes. Immediately after denaturation, slides were passed through an ethanol series (70%, 70%, 95%) at -20°C for 10 seconds each, and again air-dried.

Chromosomal proteins were digested with 100 µl of proteinase K (0.06 µg/ml in 20 mM Tris-HCl pH 7.5, 2 mM $CaCl_2$) applied to the slides under a coverslip and incubation at 37°C for 8 minutes. Coverslips were then removed and the slides passed through the same ethanol series as above, again for 10 seconds each, and air dried.

A 100 µl aliquot of denatured probe was applied to each slide and covered with a coverslip. Slides were incubated overnight at 37°C in a moist chamber, sealed with parafilm and with the slides raised off the floor of the chamber by resting the slides on glass rods spanning the diameter of the chamber.

2.4.4. IMMUNODETECTION

Following hybridization, slides were washed for 10 minutes in 1 liter of 50% formamide/2x SSC (pH 7.0) at 45°C with constant agitation. The slides were then washed twice in 1 liter of 2x SSC (pH 7.0) for 10 minutes with constant agitation. Slides were placed in BN buffer (100 mM $NaHCO_3$, 0.5% NP-40 (pH 8.0)) until ready to proceed with the immunodetection.

One hundred microliters of FAD (fluorescein avidin DCS (Vector) 5 µg/ml in BN buffer with 5% non-fat dry milk and 0.02% sodium azide) was applied to each slide, covered with a coverslip, and slides incubated in a moist chamber at 37°C for 30 minutes. Slides were washed in 1 liter of BN buffer at 45°C with constant shaking for 10 minutes, and then treated with 100 µl of BAAD (biotinylated goat anti-avidin (Vector) 12 µg/ml in BN buffer with 5% goat serum (Vector) and 0.02% sodium azide) under a coverslip in a moist chamber for 30 minutes. Slides were washed in BN buffer as above and treated with another layer of FAD as above. Slides were once again washed in BN buffer as previously described, and mounted in antifade solution (5 ml p-phenylenediamine (10 mg/ml in PBS), 45 ml glycerol pH'd to 8.0 with 0.5 M sodium carbonate-bicarbonate buffer (19.2 ml of 84 mg/ml $NaHCO_3$, 1.94 ml of 106 mg/ml Na_2CO3 made to a final volume of 50 ml with H_2O) containing DAPI (0.8 µg/ml) and propidium iodide (PI; 0.4 µg/ml). Slides could be kept in the dark at 4°C for several days.

2.4.5. IMAGE ACQUISITION

Slides were viewed under oil with a 100x objective lens with a numerical aperture of 1.3 on a Zeiss Axiophot microscope equipped for epifluorescence. The Q-like banding pattern was observed on DAPI stained chromosomes using a single band pass filter (Chroma filter set 31000). Hybridization signal was observed using a dual band pass filter set (Chroma filter set 51004) for the simultaneous detection of fluorescein isothiocyanate labeled conjugates and the propidium iodide counter stain. Images were captured on Fugi Super HG 400 ASA film at an exposure setting of -1, and a reciprocity setting of 3. The color negatives were digitized for presentation using a Sprint Scan 35 negative scanner (Polaroid) and the Adobe Photoshop program (Adobe).

2.5. FIBER-FISH

2.5.1. PRETREATMENT OF MICROSCOPE SLIDES

To ensure DNA attachment to the glass surface of the microscope slides they were coated with both gelatin and the positively charged polymer, poly-L-lysine according to the method of Heiskanen et al. (1994). Prior to coating, slides were successively immersed in 0.2M HCl, distilled water and acetone for 30 seconds each. The slides were then air-dried at room temperature, dipped in 0.15% gelatin for 5 minutes, and air-dried overnight at room temperature. Slides were submerged twice in 0.2% poly-L-lysine (Sigma) for 10 minutes each, rinsed briefly with distilled water for 30 seconds and air-dried overnight at room temperature. Slides treated as described were stored at 4°C until required. On occasion commercially coated slides were used (PROBE ON, Fisher).

2.5.2. PREPARATION OF GENOMIC DNA TARGET FIBERS

The method outlined by Heiskanen et al. (1994; 1996) was used to prepare target DNA fibers for FISH. A section equivalent to approximately 1/9 of a 200 µl pulsed field gel electrophoresis block, prepared as described above, was placed near one end of a treated microscope slide in 15 µl sterile distilled water. The slide was then held over a hot plate until the agarose block had melted completely. The resulting liquid drop, containing large DNA fibers, was extended over the surface of the slide using the end of another slide. Slides were air-dried for at least 30 minutes prior to their use in an *in situ* hybridization experiment. No fixation of the target DNA was done prior to denaturation. Slides were always processed on the same day the target fibers were prepared.

2.5.3. PREPARATION OF P1 DNA TARGET FIBERS

P1 clones (10 µg) were linearized by digestion with Not I (p2860 and p2862), or Sfi I (p2861) in a final volume of 50 µl according to suppliers directions. Digested DNAs were mixed with an equal volume 1% low melt agarose (BRL) prepared in 0.5x TBE and loaded into the wells of a 1% low melt agarose pulsed field gel. DNAs were resolved in a CHEF DR III (Biorad) apparatus in 0.5x TBE running buffer at 14°C, 6 V/cm, a reorientation angle of 120°, and a linearly ramped switch time from 1 to 6 seconds for 24 hours. Following electrophoresis, gels were stained with ethidium bromide (0.6 µg/ml) and the linearized P1 DNAs were excised from the gel after visualization with long wave UV light. The gel slices were stored at 4°C in TE until required. Target fibers for FISH were then prepared as descibed above (2.5.2.).

2.5.4. PREPARATION OF PROBES

Probes to be used in Fiber-FISH were labelled by nick translation in a final volume 50 µl. The labeling rections contained: 1 µg CsCl purified plasmid DNA, 50 mM Tris-HCl (pH 7.8), 5 mM $MgCl_2$, 10 mM β-mercaptoethanol, 10 µg/ml bovine serum albumin, 25 µM dATP, 25 µM dCTP, 25 µM dGTP, 25 µM digoxigenin-11-dUTP or 25 µM biotin-16-dUTP, and 5 µl of DNase I/DNA polymerase I mix (BRL). The reactions were allowed to proceed for 90 minutes at 14°C before being stopped by the addition of 5 µl of 10x stop buffer (0.25 M EDTA (pH 8.0), 0.5% SDS) and incubation at 37°C for 5minutes.

Probes were purified as described in 2.4.2. Probe DNA pellets were resuspended in hybridization buffer (50% formamide, 2x SSC, 10% dextran sulfate, 40 mM sodium phosphate, 0.1% SDS, 1x Denhardt's, and 100 µg/ml sonicated salmon sperm DNA (pH 7.0)) at a final concentration of 2 ng/µl. Probes were denatured at 70°C for 5 minutes and cooled on ice before use.

2.5.5. *IN SITU* HYBRIDIZATION

A 200 µl aliquot of RNase A (100 µg/ml; Sigma) in 2x SSC (pH 7.0) was placed on each slide to be hybridized and covered with a coverslip. Slides were incubated at 37°C for 1 hour in a moist chamber. Coverslips were removed and the DNA was denatured in 70% formamide/2x SSC, pH 7.0 at 70°C for 3 minutes. Slides were immediately passed through an ethanol series (70%, 70%, 95%), at -20°C, for 1 minute each, and air-dried completely. A 100 µl aliquot of the denatured probe solution was applied to each slide and a coverslip was positioned over the slide. Hybridization was allowed to proceed overnight at 37°C in a moist chamber, with the slides raised off the floor of the chamber by resting them on glass rods spanning the diameter of the chamber.

2.5.6. IMMUNODETECTION

Following hybridization, the slides were removed from the moist chamber and washed twice for 4 minutes each in 50% formamide/2x SSC, pH 7.0 and twice for 4 minutes in 2x SSC, pH 7.0, all at 45°C. Slides were then rinsed briefly in TNT buffer (100 mM Tris-HCl, pH 7.5; 150 mM NaCl; 0.05% Tween 20). Blocking was carried out by applying 100 µl of TNB buffer (100 mM Tris-HCl, pH 7.5; 150 mM NaCl; 0.5% blocking reagent (Boehringer)) under a coverslip and incubating in a moist chamber at 37°C for 30 minutes.

Following the blocking step, coverslips were removed and 100 µl of antibody solution I (5 µg/ml fluorescein-avidin DCS (Vector), 5 µg/ml mouse

monoclonal anti-digoxigenin (Boehringer) in TNB buffer) was added under a coverslip and incubated in a moist chamber at 37°C for 30 minutes. Slides were washed three times for 5 minutes each in TNT buffer at room temperature. Slides were then treated with 100 µl of antibody solution II (5 µg/ml biotinylated goat anti-avidin D (Vector), 10 µg/ml digoxigeninylated sheep anti-mouse Ig-F(ab')$_2$-fragment (Boehringer) in TNB buffer) under a coverslip and incubated in a moist chamber at 37°C for 30 minutes. The slides were washed as before in three changes of TNT buffer. Finally, 100 µl of antibody solution III (5 µg/ml fluorescein-avidin DCS (Vector), 12.5 µg/ml rhodamine-sheep anti-digoxigenin (Boehringer) in TNB buffer) was applied under a coverslip and incubated at 37°C in a moist chamber for 30 minutes. The slides were washed again as previously described, excess moisture blotted from the slides, and 35 µl of antifade mounting solution without DAPI and PI applied under a coverslip. Slides were analyzed immediately, or kept in the dark at 4°C for several days before the analysis was completed.

2.5.7. IMAGE ACQUISITION

Slides were examined as described previously (2.4.5), however. hybridization signals were observed using a triple band pass filter set (Chroma set 61002) for the simultaneous detection of DAPI signal as well as fluorescein isothiocyanate and trimethyl rhodamine isothiocyanate labeled conjugates, and images were captured on Fugi Super HG 1600 ASA film at an exposure setting of 0 or-1, and a reciprocity setting of 4.

2.5.8. IMAGE ANALYSIS

Images on negatives from prints demonstrating suitable hybridization to the linearized P1 DNA targets were projected on to a sheet of paper using a slide projector. Measurements for the purposes of calculating the position of alpha-2u-globulin genes on each of the P1 clones were made from the projected image. Interval sizes on each P1 molecule examined were determined in micrometers, percent of the entire length of the P1 molecule, and kilobases. Interval sizes in kilobases were determined by multiplying the known length of the P1, as determined by pulsed field gel electrophoresis, by the percent of the length represented by the interval under investigation. Similarily, interval sizes in kilobases could be determined by dividing the particular interval size in micrometers by the degree of stretching calculated for the P1 molecule being investigated. The degree of stretching for each P1 molecule was determined by dividing the length of the P1 molecule (in micrometers) by its size (in kilobases), as determined by pulsed field gel electrophoresis. The size of gaps between genes was determined by measurements made from the end of one signal to the start of the next signal. To determine the distance between terminal genes and the end of the

linearized P1 DNAs, measurements were made from the end of the hybridization signal produced by hybridization to P1 DNA, to the beginning of the signal resulting from hybridization to 91R91. The size of the genes was determined by measuring the length of the signal produced by hybridization to 91R91.

The same method of projecting images from a print negative on to a sheet of paper was used in the analysis of images obtained following FISH to rat genomic DNA.

CHAPTER 3.
RESULTS
MOLECULAR ANALYSES

3.1. ESTIMATION OF ALPHA-2U-GLOBULIN GENE NUMBERS

3.1.1. SOUTHERN BLOT ANALYSIS

To estimate the number of genes in the alpha-2u-globulin family, and to determine whether classes of genes could be identified, genomic Southern blots were hybridized with probes derived from alpha-2u-globulin gene 91.

Analysis of the 5′ region of the alpha-2u-globulin gene was performed using the 1.2 kb Hind III-Eco RI fragment from the 5′ region (-700 to +500) of gene 91 (Figure 2.1) as a probe, and blots washed using either low (4X SSPE, 65°C) (Figure 3.1), or high stringency (0.1X SSPE, 65°C).

The autoradiographs generated following differential washing of blots hybridized with the 1.2 kb Hind III-Eco RI probe display a similar, complex pattern of bands, with no differences in relative band intensities. This indicates that the members of the alpha-2u-globulin genes are highly conserved at their 5′ ends and that it is unlikely that classes of alpha-2u-globulin genes that are significantly divergent in the 5′ region of the gene exist. At least 9 bands are apparent in the lane with Eco RI digested DNA. The bands differ in intensity, indicating that some of the bands represent more than one alpha-2u-globulin gene which share similar sized Eco RI fragments.

A similar analysis using the 600 bp Ava I-Bgl II fragment, and the adjacent 500 bp Bgl II-Hind III fragment from the 3′ end of gene 91 (Figure 2.1) was performed to determine whether classes of alpha-2u-globulin genes differing in their 3′ ends could be detected (Figure 3.2). Kurtz (1981) reported the isolation of two alpha-2u-globulin genomic clones whose 3′ ends did not hybridize with the 500 bp Bgl II-Hind III fragment from gene 91. The present study did not indicate significant differences between the hybridization patterns obtained with the two probes from the 3′ end of alpha-2u-globulin gene 91. This suggests that genes with 3′ ends divergent from gene 91 are certainly in the minority. The conclusion that the majority of genes share homology in their 3′ ends is certainly substantiated by the sequences of the 12 genes reported in Figure 2.3. Genomic DNA digested with Eco RI and

Figure 3.1. Southern hybridization analysis of rat genomic DNA. Rat genomic DNA digested with Eco RI, E (lane 1); Dra III, D (lane 2); Bst XI, Bs (lane 3); and Bgl I, Bg (lane 4) and hybridized with the 1.2 kb Hind III-Eco RI probe from the 5′ end of gene 91. The blot was washed in 4X SSPE at 65°C. Alpha-2u-globulin gene 91 from Figure 2.1 is shown to indicate the origin of the probe used in the hybridization.

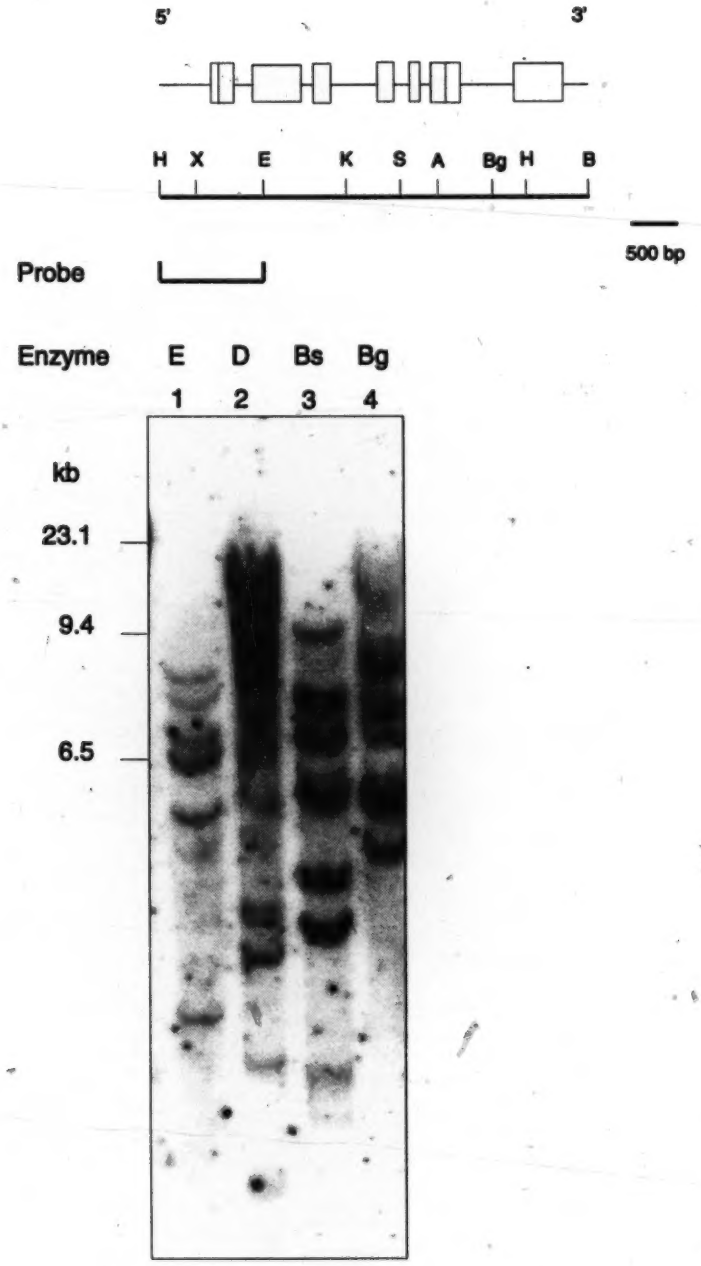
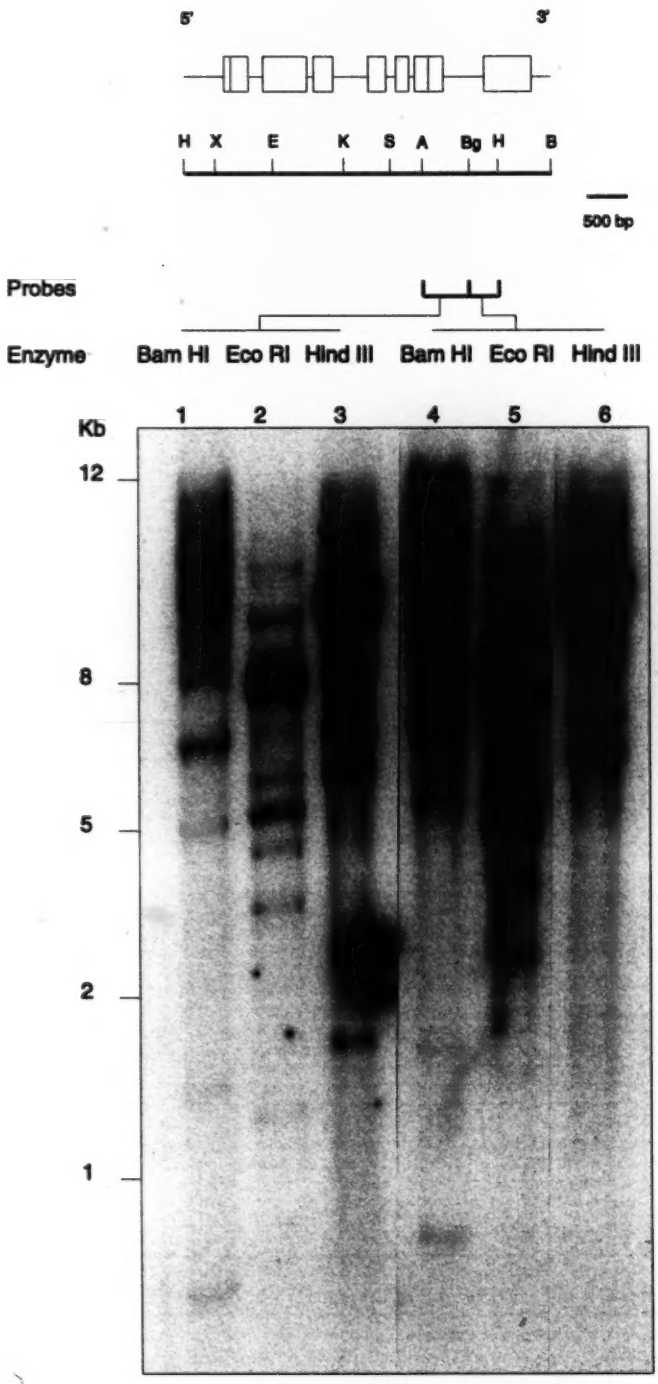

5'
3'
H X E K S A Bg H B
500 bp
Probe
Enzyme E D Bs Bg
1 2 3 4
kb
23.1
9.4
6.5

Figure 3.2. Analysis for divergence in the 3′ ends of alpha-2u-globulin genes. Rat genomic DNA was digested with Bam HI (lanes 1, 4), Eco RI (lanes 2, 5), or Hind III (lanes 3, 6) and hybridized with the 500 bp Bgl II-Hind III fragment from 91R91 (lanes 1-3), or the 600 bp Ava I-Bgl II fragment from 91R91 (lanes 4-6). The blots were washed in 1X SSPE at 65°C. Alpha-2u-globulin gene 91 from Figure 2.1 is shown to indicate the origin of the probes used in the hybridization.


5'
3'
H X E K S A Bg H B
500 bp
Probes
Enzyme
Bam HI Eco RI Hind III Bam HI Eco RI Hind III
Kb
1 2 3 4 5 6
12
8
5
2
1

hybridized with the probes from the 3′ end, displayed a banding pattern equally as complex as that observed after hybridization with the probe from the 5′ end of gene 91. At least 12 bands, demonstrating differences in relative band intensities, were detected, indicating that there are more than 12 alpha-2u-globulin genes in the rat genome. An estimate 20-22 genes, would more adequately reflect both the number and intensity of bands observed on the auto-radiograph.

Hybridization of pSK2-3 (Figure 2.1), which contains sequences from +700 to +3950 of gene 91, to digested rat genomic DNA also gave a complex pattern of bands on autoradiographs (Figure 3.3). This indicates, that in addition to the highly conserved coding sequences, the upstream regions at least as far as 4 kb are also conserved among many members of the alpha-2u-globulin family. This conclusion is substantiated by restriction mapping results (Kurtz, 1981; Wang, 1996) and results from the differential stringency screening of a rat genomic library (3.2.2). Furthermore, this probe was found to weakly hybridize to discrete fragments in restriction enzyme digested and Southern blotted mouse genomic DNA, indicating the possible conservation of these sequences upstream of the mouse MUP genes.

3.1.2. SLOT BLOT ANALYSIS

A slot blot prepared with various quantities of plasmid 91R91 representing 5-50 alpha-2u-globulin gene equivalents, and replicate samples of rat genomic DNA was hybridized with the 4 kb Hind III fragment of plasmid 91R91. The resulting autoradiogram (Figure 3.4) was subject to a densitometric scan to determine the mean optical densities of each slot on the autoradiogram. Graphical analysis (Figure 3.5) of the results presented in Table 3.1 indicates that there are approximately 20 alpha-2u-globulin genes in the rat genome. This estimate is consistent with the estimate of 18-20 made by Kurtz (1981) based on solution hybridization kinetics, and the number of isoforms observed following isoelectric focussing (Gubits et al., 1984).

3.2. LAMBDA CLONE CHARACTERIZATION

3.2.1. HOMOLOGY MAPPING OF LAMBDA GENOMIC SUBCLONES

The extent of homology between the upstream region of the alpha-2u-globulin gene in subclone pL1 and that of the upstream region of the gene represented by clones pL2, pL8, pL9, and pL10 was investigated by examining the ability of probes derived from pL1 to hybridize to lambda subclones pL2 and pL10, which contain the most downstream and upstream

Figure 3.3. Analysis of the 5′ upstream region of alpha-2u-globulin genes. Rat genomic DNA was digested with Eco RI (lane 2), or Bam HI (lane 3). Mouse genomic DNA was digested with Eco RI (lane1). All lanes were hybridized with pSK2-3 containing the upstream region of gene 91. The blot was washed in 1X SSPE at 65°C. Alpha-2u-globulin gene 91 from Figure 2.1 is shown to indicate the origin of the probe used in the hybridization.

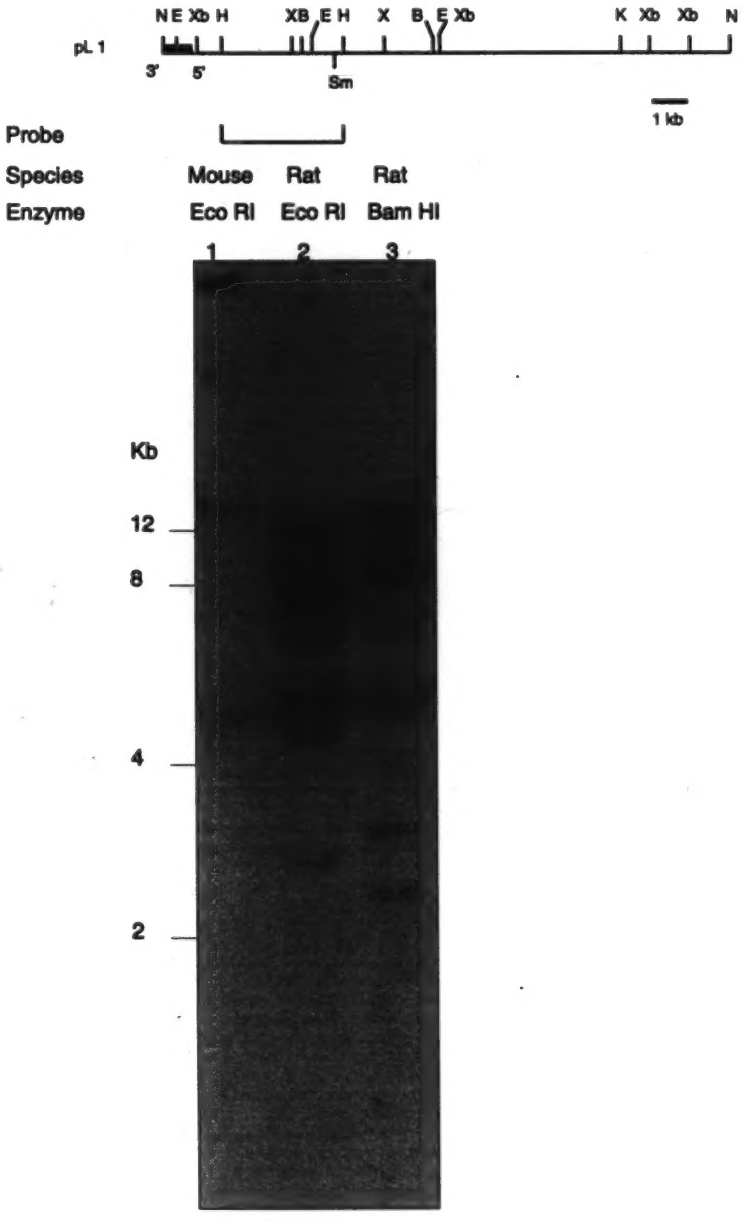

N E Xb H
XB E H
X
B E Xb
K Xb Xb N
pL 1
3'
5'
Sm
1 kb
Probe
Species
Enzyme
Mouse
Rat
Rat
Eco RI
Eco RI
Bam HI
1
2
3
Kb
12
8
4
2

Figure 3.4. Alpha-2u-globulin quantitative slot blot. The indicated number of alpha-2u-globulin gene equivalents were blotted in the top row of the blot. Twenty micrograms of rat genomic DNA was blotted in triplicate in the second row of the blot. The blot was hybridized to labelled 91R91 and washed in 1X SSPE at 65°C.

Table 3.1. Densitometric Analysis of Alpha-2u-Globulin Quantitative Slot Blot

Sample	Mean Optical Density	Average Optical Density
Plasmid 91R91		
50 copies	0.303	
30 copies	0.270	
28 copies	0.266	
26 copies	0.252	
24 copies	0.258	
22 copies	0.250	
20 copies	0.223	
18 copies	0.202	
15 copies	0.183	
10 copies	0.159	
5 copies	0.144	
Rat genomic DNA		
sample 1	0.225	0.227
sample 2	0.239	
sample 3	0.219	

The mean optical denities are provided for the hybridization signals generated by hybridization of the 4 kb Hind III fragment from 91R91 to each of the indicated samples. An average optical density is shown for the rat genomic DNA, which was blotted in triplicate.

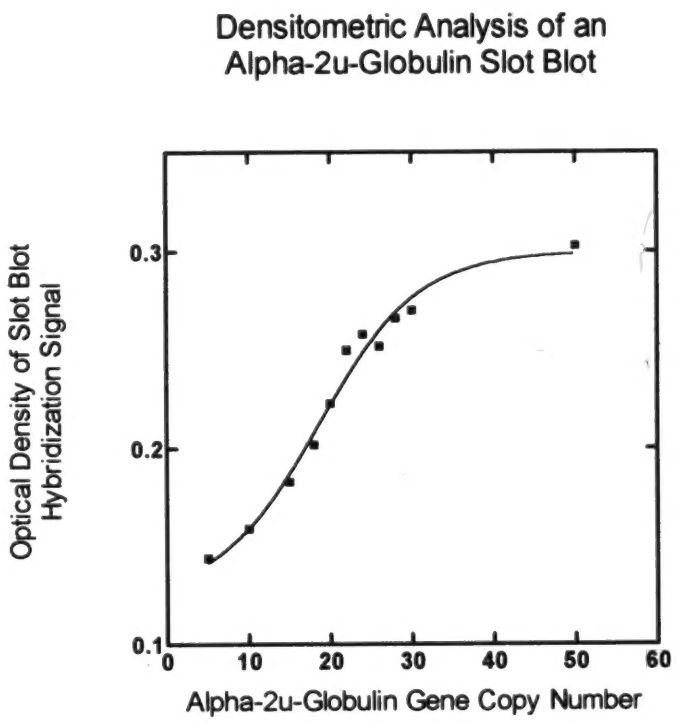


Figure 3.5. Graphical results from an alpha-2u-globulin quantitative slot blot. The optical density of the slot blot hybridization signal was plotted against the copy number blotted in each slot using the. Boltzmann sigmoidal function. The average optical density for the slots containing rat genomic DNA was 0.227, which corresponds to an alpha-2u-globulin gene copy number of approximately 20. Similar estimates were arrived at following analysis with other mapping functions.

sequences respectively from the gene represented by these four subclones. The results of the hybridization analysis are schematically represented in Figure 3.6 (A). Probes 2 and 3, which are prepared from sequences between +700 to +4300 bp of subclone pL1 were found to cross-hybridize to regions of each of the other four subclones (only pL2 and pL10 shown in Figure 3.6 (A)). Probe 4, derived from sequences between +4.3 to +5.7 kb of subclone pL1, hybridized to sequences at corresponding upstream positions of clones pL8, pL9, and pL10; clone pL2 does not extend that far in the 5' direction, explaining the failure of probe 4 to hybridize to this clone. Probe 8, which extend from +6.6 to +11.4 kb of pL1 was found to hybridize with the corresponding upstream region of pL10. Probes 9 and 7, which extend from +12.1-+13.3 and +13.3-+15.6 of pL1 respectively did not hybridize to the other subclones. This result indicates that the extensive homology that has been noted between the coding regions of alpha-2u-globulin genes (see Figure 2.3) also extends further than 6.6 kb upstream, at least in the case of the two genes represented by these clones. The extensive homology between alpha-2u-globulin genes was also demonstrated by screening the same rat genomic library from which pL1, 2, 8, 9, and 10 were obtained from with probes 3 and 8 (see 3.2.3).

In the course of the above analysis, a somewhat surprising observation was made. Probes 4 and 8 which were isolated from the upstream sequences of the gene in clone pL1 were found to hybridize not only to the upstream sequences of the alpha-2u-globulin gene represented by clones pL8, 9, and 10, but also to sequences downstream of alpha-2u-globulin in clones pL2, 8, and 9. The order in which sequences that hybridize with probes 4 and 8 appear in the downstream region of clone pL2, which has the most extensive 3′ region, is inverted with respect to the order that they appear in the upstream region. This makes it unlikely that the sequences present in the downstream region of clone pL2 represent the upstream flanking sequences of the adjacent alpha-2u-globulin gene.

The cross-hybridizing regions from the 5′ flanking sequences of pL1 and 3′ flanking sequences of pL2 were examined in more detail by restriction endonuclease analysis of pL1 subclone pL1-1, and pL2 subclone pL2-1 (Figure 3.6 (B)). These two clones contain the regions hybridizing to probes 4 and 8 from the 5′ region of pL1 and the 3′ region of pL2 respectively. Restriction mapping demonstrated that the two cross-hybridizing segments represent a region of inverted symmetry which flanks at least these two alpha-2u-globulin genes.

3.2.2. DIFFERENTIAL STRINGENCY SCREENING OF A LAMBDA GENOMIC LIBRARY

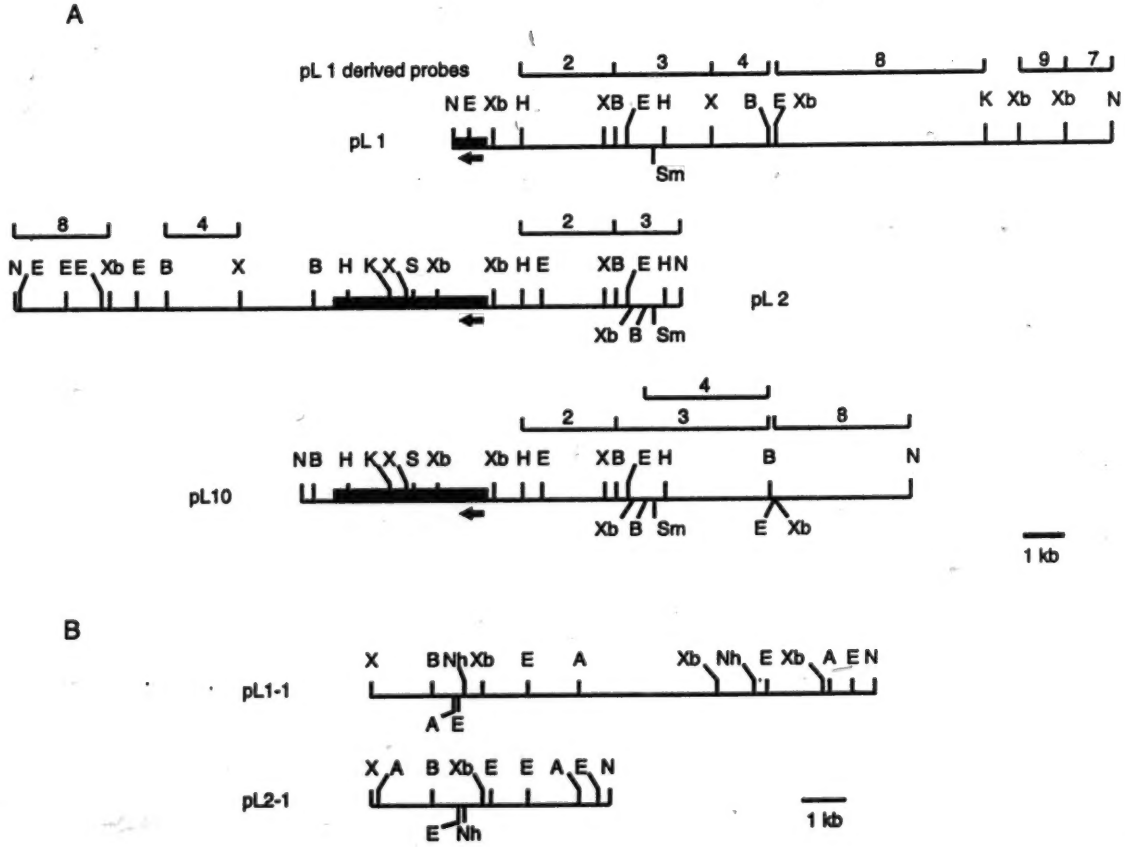
The possible existance of two alpha-2u-globulin gene classes, similar to

Figure 3.6. Genomic subclone homology mapping.

(A) Restriction maps of pL1, pL2, and pL3. Probes derived from pL1, which contains the upstream region of alpha-2u-globulin gene 91, are indicated above the restriction map for pL1. The limits of hybridization of the pL1 derived probes on each of the clones pL2 and pL10 are indicated above the respective clones. The position of the gene in these clones is shown by the colored bar.

(B) Detailed restriction maps of the terminal subclones from the right end of pL1 (pL1-1), and the left end of pL2 (pL2-1).

Restriction sites shown in both panels: A, Acc I; B, Bam HI; E, Eco RI; H, Hind III; K, Kpn I; N, Not I; Nh, Nhe I; S, Sal I; Sm, Sma I; X, Xho I; Xb, Xba I. The Not I restriction sites are those of pBluescript.


A
pL 1 derived probes
2
3
4
8
9
7
N E Xb H
XB E H
X
B E Xb
K Xb Xb N
pL 1
Sm
8
4
2
3
N E EE Xb E B
X
B H K X S Xb
Xb H E
XB E H N
pL 2
Xb B Sm
4
2
3
8
NB H K X S Xb
Xb H E
XB E H
B
N
pL10
Xb B Sm
E Xb
1 kb
B
X
B Nh Xb
E
A
Xb Nh E Xb A E N
pL1-1
A E
X A B Xb E E A E N
pL2-1
E Nh
1 kb

those identified for the mouse MUP genes by differential hybridization, was investigated using a differential stringency screen of a rat lambda genomic library. A duplicate set of six filters representing 1.5 rat genome equivalents (each filter contained approximately 50 000 plaques) was hybridized to the 4 kb Hind III from plasmid 91R91. One set was washed under conditions of low stringency and a total of 37 plaques were found. The low stringency wash conditions (4X SSC, 42°C) were such as to permit detection of hybrids with sequences with a mismatch of approximately 30% to the probe. The other set of filters, hybridized with the same probe, but washed under high stringency conditions (0.1X SSC, 65°C), allowing the detection of sequences with about a 5% mismatch with the probe. This filter set was positive for the same 37 plaques identified under the low stringency washing conditions. This result indicates that classes of alpha-2u-globulin genes, recognized by differential hybridization conditions similar to those which reveal classes of MUP genes, do not exist. Should classes of alpha-2u-globulin genes exist, they would have to vary from each other by less than 5% sequence difference.

3.2.3. HOMOLOGOUS UPSTREAM SEQUENCES AMONG ALPHA-2U-GLOBULIN GENES

The results obtained from the analysis of the two alpha-2u-globulin genes represented by the 5 lambda genomic subclones pL1,2, 8, 9, 10 indicated that there was extensive homology between the 5′ upstream regions of the alpha-2u-globulin genes (see 3.2.1). To further examine the possibility that extensive 5′ homology is a common feature of the alpha-2u-globulins, the duplicate filter sets previously screened with the 4 kb Hind III fragment of 91R91 (see 3.2.2) were screened with two probes, 3 and 8, derived from pL1 (see Figure 3.6 (A)). The results of this screen are provided in Table 3.2 and are combined with the results obtained when the same filter set was hybridized to the 4 kb Hind III fragment from 91R91 (3.2.2).

Of the 37 plaques previously identified with the 4 kb Hind III fragment of 91R91, 14 were also positive for both probes 3 and 8, 15 were positive for probe 3, and 1 was positive for probe 8 alone. The remaining 7 plaques that hybridized to the 4 kb Hind III fragment from 91R91 were not positive for either probe 3, or 8. The estimated maximum insert size for a clone positive only for the 4 kb Hind III fragment of 91R91 would be about 11.5 kb. A fragment of this size would easily be accommodated by the lambda DASH II cloning vector used in the construction of this library, which is reported to accept inserts ranging in size from 9 to 23 kb.

In addition to the positive plaques mentioned above, 7 plaques were found to be positive for probes 3 and 8, but not the 4 kb Hind III fragment of 91R91. An additional 19 plaques hybridized to probe 8 only, and 5 plaques

Table 3.2. Analysis of a Rat Genomic Lambda Library

Probe(s)	Positive Plaques
4 kb Hind III fragment from 91R91	2F 4G 5A 5C 6A 6B 6D
Probe 3	1E 2O 3G 3H 4N
Probe 8	1F 1G 1H 1I 2K 2L 2M 2N 3K 3L 3M 4K 4L 4M 5E 5F 5G 5H 5I
4 kb Hind III fragment from 91R91/ Probe 3/Probe 8	1A 1B 2A 2B 2C 2E 2G 2I 2J 3B 3E 4C 4E 6E
4 kb Hind III fragment from 91R91/ Probe 3	1C 1D 2D 2H 3C 3D 3F 4A 4B 4D 4F 5B 6C 6G 6F
4 kb Hind III fragment from 91R91/ Probe 8	3A
Probe 3/Probe 8	3I 3J 4H 4I 4J 5D 6H
600 bp Ava I-Bgl II fragment from 91R91	1A 1B 1C 2A 2B 2C 2D 2F 2H 2I 2J 3B 3C 3E 3F 4A 4B 4C 4D 4F 4G 4H 5A 5B 5C 6A 6B 6C 6D 6F 6G
500 bp Bgl II-Hind III fragment from 91R91	1A 1C 2A 2B 2C 2F 2H 3A 3B 3C 4A 4B 4C 4D 4G 4H 5A 5B 5C 6A 6B 6C 6D 6F 6G

The first column of the table lists the probe(s) used to screen a set of six filters each containing approximately 50 000 phage from a rat genomic lambda library. The second column lists the positive plaques identified with each probe(s). Positive plaques are denoted with a number, to identify which of the 6 filters they are on, and a letter to specifically identify the plaque on the filter.

were found to hybridize to probe 3 only. This latter class is unexpected given the restriction maps of the available lambda subclones (Figure 2.2) and the reported size range of inserts for this particular library. Based on the existing maps, a phage positive for only probe 3 would have a maximum insert size of approximately 6.5 kb, making it smaller than the 9kb reported to be the smallest fragment size accepted by the vector. A possible explanation for this class of plaques could be that there are alpha-2u-globulin genes with insertions in the region between the gene and those sequences complementary to probe 3, and/or insertions in the region between the sequences complementary to probe 3 and 8 thus generating fragments that could be accommodated by the cloning vector, and that contain alpha-2u-globulin associated sequences complementary to only probe 3

The number of positive plaques observed when each of the three probes, 3, 8, and the 4 kb Hind III fragment of 91R91, is used separately to screen the same filter sets is 41, 41, and 37 respectively. These results indicate that sequences homologous to the two probes obtained from the 5′ flanking sequences of the alpha-2u-globulin gene represented in pL1 are common to a large majority, if not all of the alpha-2u-globulin genes in the rat genome. This evidence further supports the earlier observation that homology between alpha-2u-globulin genes can extend at least 6.6 kb upstream of the gene.

3.2.4. DIVERGENCE IN THE 3′ ENDS OF ALPHA-2U-GLOBULIN GENES

Although two distinct classes of alpha-2u-globulin genes could not be identified on the basis of differential hybridization to the 4 kb Hind III fragment from 91R91, it has previously been suggested that classes of alpha-2u-globulin genes with divergent 3′ non-coding regions may exist (Kurtz, 1981). Kurtz reported the isolation of 2 alpha-2u-globulin gene containing lambda genomic clones with 3′ non-coding regions evolutionarily divergent to that of 91R91.

The prevalence of alpha-2u-globulin genes with divergent 3′ regions was further investigated by screening the set of duplicate filters containing the rat genomic lambda library (see 3.2.2) with two probes derived from adjacent regions in the 3′ end of 91R91. The probes used were the 600 bp AvaI-BglII fragment and the 500 bp BglII-HindIII fragment (see Figure 2.1). The results of the screen are presented in Table 3.2.

Of the 37 plaques identified with the 4 kb Hind III fragment of 91R91, 31 were identified with the 600 bp Ava I-Bgl II probe, while 25 were identified with the more 3′ 500 bp Bgl II-Hind III probe, despite the two probes being prepared from adjacent segments of the alpha-2u-globulin gene 91.

These results are compatible with the earlier observation, made by Kurtz (1981), of divergence in the non coding 3′ region of alpha-2u-globulin genes. However, as indicated by Southern hybridization to genomic DNA (3.1.1), and confirmed by the library screen, the majority of alpha-2u-globulin genes have 3′ regions that are similar to that present in clone 91R91.

3.3. P1 CLONE CHARACTERIZATION

3.3.1. SIZE OF P1 INSERTS

Three P1 clones, identified by screening the P1 library with the PCR product obtained with the primer pair DAMF/DAMR, were obtained and have been designated p2860, p2861, and p2862. Isolation of significant quantities of P1 DNA from these clones, suitable for mapping, proved to be difficult. The difficulties were not merely those associated with the isolation of a single copy plasmid. Induction of the P1 lytic operon with IPTG to increase the copy number to approximately 20 copies/cell failed to produce the desired increase in plasmid yield. Growth in the rich medium, TB, both with and without induction with IPTG also failed to increase the yield and quality of the P1 plasmids isolated from the three strains.

The problem encountered in this study with the isolation of P1 DNA, although not widely reported by other researchers, has been encountered in attempts to isolate sequenceable P1 DNA templates (Kimmerly et al., 1994). These researchers devised a mating scheme to transfer P1 plasmids from the original host strain, NS3529, to a new host, DH10B which proved to be a superior host strain for the preparation of sequencing templates. Electroporative transfer of p2860, p2861, and p2862 to DH10B allowed for the isolation of about 200 μg of high quality plasmid DNA from 500 ml culture volumes.

Two reasons can be suggested for the improved quality and quantity of P1 plasmid DNA from DH10B. First, DH10B, unlike NS3529, does not express the P1 recombinase. The constitutive expression of the recombinase in the original host, although required initially to generate the P1 plasmid, would tend to result in the production of P1 multimers by recombination between lox sites on different P1 molecules. It is possible that these multimers may not be released efficiently from genomic DNA and cellular debris following cell lysis. Second, DH10B carries a mutation in the *end A* gene. The *end A* gene codes for an endonuclease that has been associated with poor plasmid recovery (Taylor et al., 1993).

The size of the genomic insert carried by each of the three P1 clones was determined by restriction endonuclease digestion with Not I, Sfi I, and Sal I

(Figure 3.7). The three clones; p2860, p2861, and p2862 contain inserts of 75, 78, and 87 kb respectively. Figure 3.7 demonstrates that clones p2860 and p2862 lack restriction sites for Not I and Sfi I in the insert. Clone p2861 contains a Not I restriction site in the insert as evidenced by a size difference of about 3 kb between the digestion products with Sfi I and Not I. Clones p2860 and p2862 each appear to possess four Sal I sites, while p2861 has three. These three genomic clones are the largest segments of DNA isolated to date that contain alpha-2u-globulin sequences.

3.3.2. P1 GENE CONTENT

To assess the alpha-2u-globulin gene content of the P1 clones, the pulsed field gel shown in Figure 3.7 was blotted and hybridized with the 1.2 kb Hind III-Eco RI fragment from 91R91 (Figure 3.7). The lanes containing Sal I digested DNA were informative. Three of the fragments generated following cleavage of p2860 with Sal I hybridized to the probe, indicating that the 74 kb genomic insert in p2860 carries 3 alpha-2u-globulin genes. One Sal I fragment produced following digestion of p2861 hybridized with the 1.2 kb Hind III-Eco RI probe from gene 91, suggesting a single gene on the insert of this clone. P1 clone p2862 contains two alpha-2u-globulin genes, as suggested by hybridization to two Sal I fragments.

The gene content of the three P1 clones as suggested by the Sal I digestions was examined further by digesting the three clones with Bam HI and Eco RI and hybridization to the 1.2 kb Eco RI-Sal I fragment from 91R91 (Figure 3.8). The results reinforce the earlier observations from the hybridization to Sal I digested DNA. It is concluded, based on the results obtained using three different enzymes that p2860 contains three alpha-2u-globulin genes, p2861 one gene, and p2862 has two genes.

The restriction fragment length heterogeneity observed when digested total rat genomic DNA is hybridized with various gene probes (Figures 3.1-3.3) is demonstrated once again in the analysis of the 3 P1 clones. These results indicate that although it is apparent from genomic Southerns that many alpha-2u-globulin genes are present on similarly sized restriction fragments, genes on similar sized fragments are not necessarily clustered together, although it appears that the two alpha-2u-globulin genes on p2862, are on almost identically sized Bam HI restriction fragments.

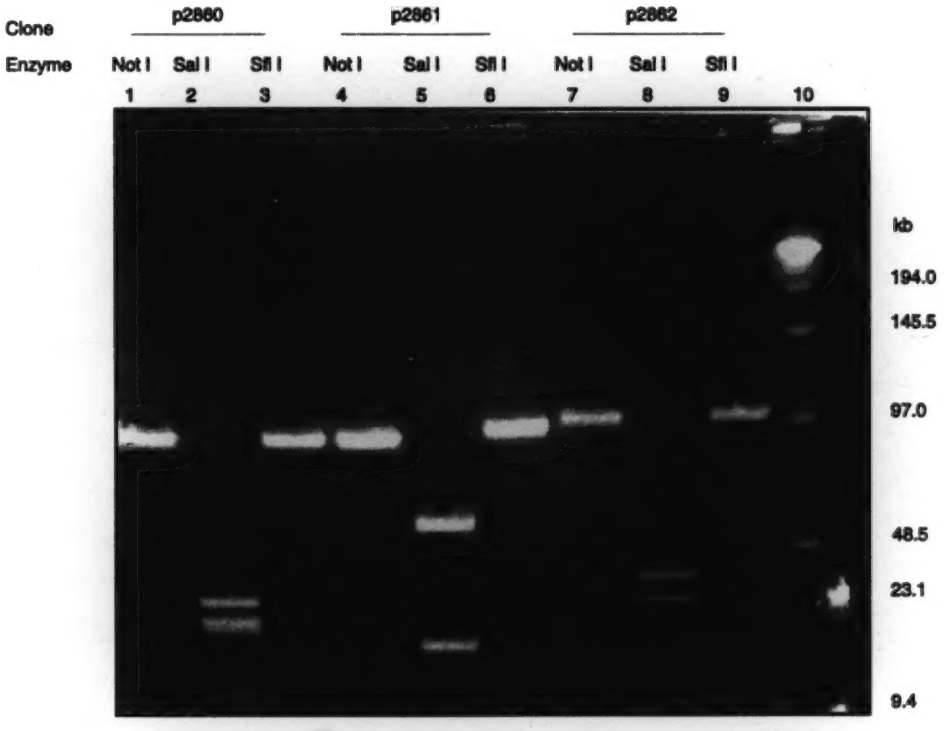
Length heterogeneity was also observed in restriction fragments from the three P1 clones hybridized with the 1.2 kb Hind III-Eco RI fragment from 91R91 (Figure 3.9). Again this would suggest that adjacent genes, and gene regions from adjacent genes are not necessarily present on identically sized restriction fragments.

Figure 3.7 Analysis of P1 clones.

(A) Pulsed field gel electrophoretic separation of restriction endonuclease digested P1 clone DNA. Lanes 1-3, p2860; lanes 4-6, p2861; lanes 7-9, p2861; lane 10, lambda concatamer and lambda Hind III. Restriction enzymes used: lanes 1, 4, 7, Not I, N; lanes 2, 5, 8, Sal I, S; lanes 3, 6, 9, Sfi I, Sf. Electrophoresis conditions: 6 V/cm, 12 hours, ramped switch times from 0.5-2 seconds, 14°C.

(B) Same gel as in (A), blotted and hybridized with the 1.2 kb Hind III-Eco RI fragment from the 5' end of 91R91. The arrows in both panels indicate doublets.

A


Clone
p2860
p2861
p2862
Enzyme
Not I
Sal I
Sfi I
Not I
Sal I
Sfi I
Not I
Sal I
Sfi I
1
2
3
4
5
6
7
8
9
10
kb
194.0
145.5
97.0
48.5
23.1
9.4

B

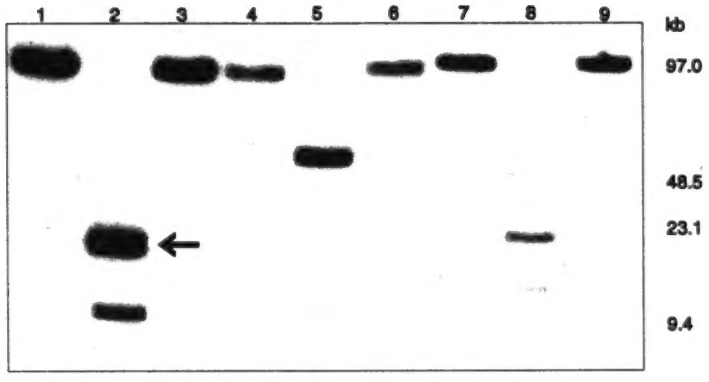

1
2
3
4
5
6
7
8
9
kb
97.0
48.5
23.1
9.4

Figure 3.8. P1 clone alpha-2u-globulin gene content. Southern hybridization analysis of Bam HI digested (lanes 1, 3, 5) and Eco RI digested (lanes 2, 4, 6) p2860 (lanes 1, 2), p2861 (lanes 3, 4), and p2862 (lanes 5, 6) hybridized with the 1.2 kb Eco RI-Sal I fragment of 91R91. Alpha-2u-globulin gene 91 from Figure 2.1 is shown to indicate the origin of the probe used in the hybridization.

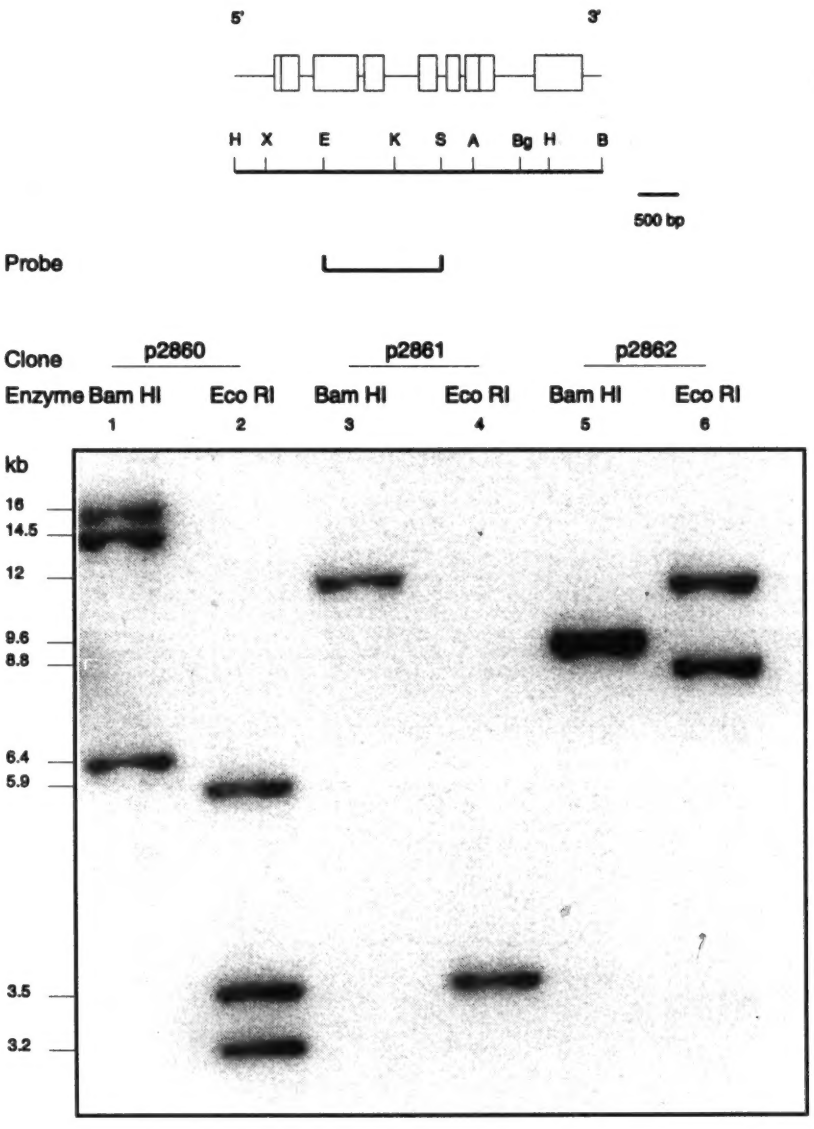
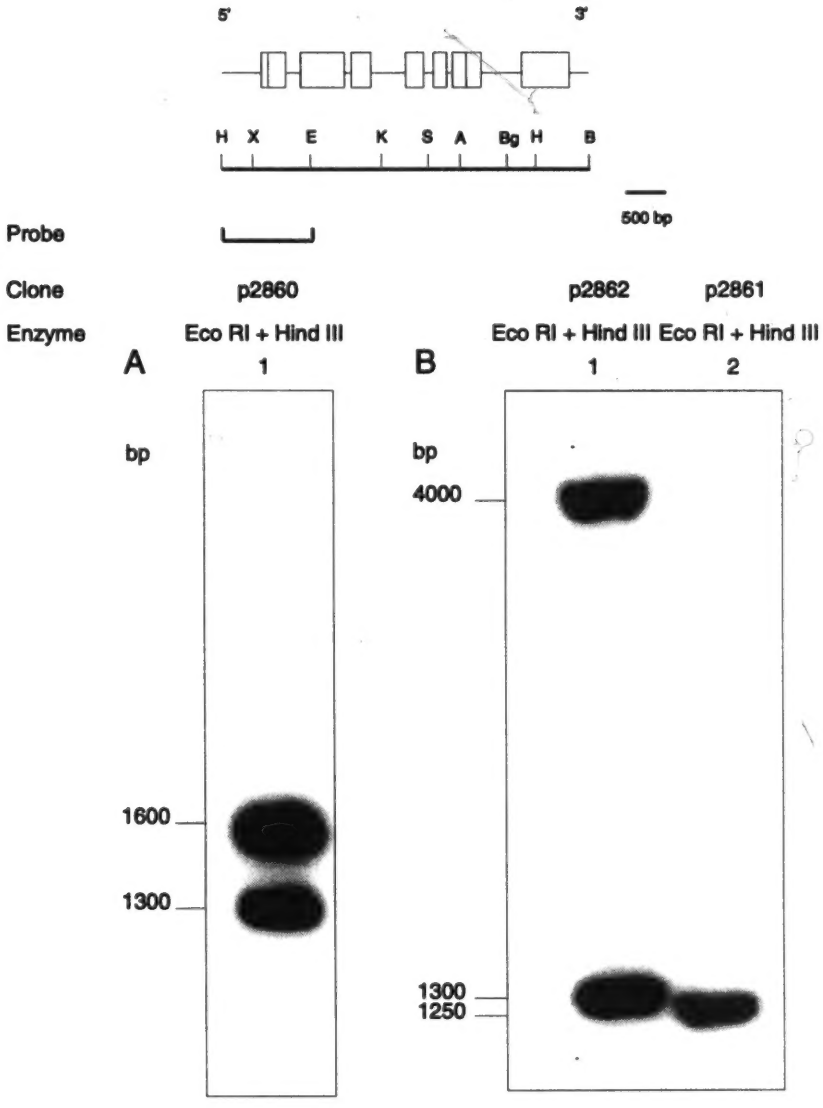

5'
3'
H X E K S A Bg H B
500 bp
Probe
Clone
p2860
p2861
p2862
Enzyme Bam HI Eco RI Bam HI Eco RI Bam HI Eco RI
1 2 3 4 5 6
kb
16
14.5
12
9.6
8.8
6.4
5.9
3.5
3.2

Figure 3.9. Restriction site polymorphism in the upstream region of alpha-2u-globulin genes.

(A) Southern hybridization analysis of Eco RI-Hind III digested p2860 hybridized with the 1.2 kb Eco RI-Hind III probe from 91R91.

(B) Southern hybridization of Eco RI-Hind III digested p2861 (lane 1) and p2862 (lane 2) probed as in (A)

Alpha-2u-globulin gene 91 from Figure 2.1 is shown to indicate the origin of the probe used in the hybridizations.


5'
3'
H X E K S A Bg H B
500 bp
Probe
Clone
p2860
p2862
p2861
Enzyme
Eco RI + Hind III
Eco RI + Hind III
Eco RI + Hind III
A
1
bp
1600
1300
B
1
2
bp
4000
1300
1250

3.3.3. P1 RESTRICTION MAPS

A physical map of the cloned genomic segments carried by the P1 clones was generated by mapping sites for Not I, Nru I, Pme I, Sac I, Sal I, and Sma I. The maps for the three clones are presented in Figure 3.10. The position(s) and orientation of alpha-2u-globulin genes on each of the three P1 clones as determined by Southern hybridization is also indicated in Figure 3.10. The three P1 clones, based on restriction site analysis, appear not to be overlapping.

3.4. PULSED FIELD GEL MAPPING

3.4.1. LONG RANGE MAP OF THE ALPHA-2U-GLOBULIN LOCUS

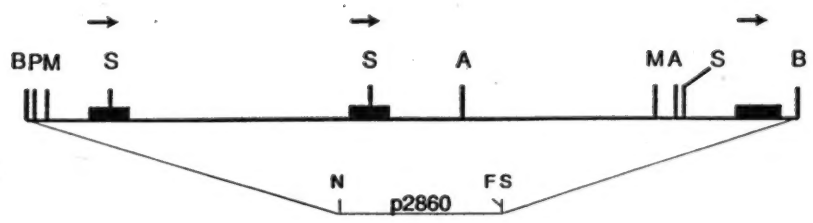
To construct a long range restriction map of the alpha-2u-globulin locus, agarose gel blocks containing restriction endonuclease digested rat genomic DNA were subjected to pulsed field gel electrophoresis. A variety of restriction enzymes (Aat II, Asc I, BssH II, Mlu I, Not I, Nru I, Pac I, Pme I, Pvu I, and Sfi I) commonly used in pulsed field gel electrophoresis were examined for their suitability for the construction of a restriction map encompassing the alpha-2u-globulin cluster.

After Southern transfer and hybridization with the 4 kb Hind III fragment from 91R91, to rat genomic DNA, the digests with Asc I, BssH II, Not I, Pvu I, or Sfi I, failed to show hybridization to any fragments below the compression zone (approximately 1 Mb under these conditions). The fact that a Not I restriction site was mapped within the P1 clone p2861 indicates that all the genes are on one Not I fragment and the next nearest Not I restriction site must be greater than 1 Mb from the NotI site mapped to p2861. Alternatively, there could be more proximal Not I restriction sites which are protected from cleavage by methylation, but no "partial cleavage" bands were observed.

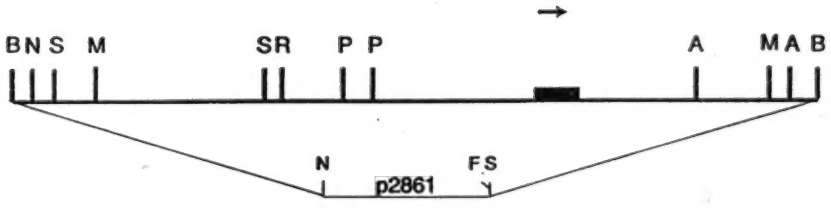
Digestion of rat genomic DNA with the enzymes Aat II, or Pac I followed by Southern transfer and hybridization with the 4 kb Hind III fragment from 91R91 produced a complex pattern of bands which were all smaller than about 75 kb. Digestion with Pme I followed by hybridization with the 4 kb Hind III fragment from 91R91 produced a pattern of bands ranging in size from about 50-300 kb (Figure 3.11). The multitude of bands indicates recognition sites for these three enzymes occur frequently throughout the cluster, making assembly of a long range restriction map based on these enzymes difficult. However, the somewhat less frequent occurrence of restriction sites for Pme I compared to both Aat II and Pac I, made it a good candidate enzyme to use in the characterization of the P1 clones.

Figure 3.10. Restriction maps of three alpha-2u-globulin containing bacteriophage P1 clones. (A) p2860. (B) p2861. (C) p2862. Restriction enzymes: A, Sst II; B, Bam HI; F, Sfi I; M, Sma I; N, Not I; P, Pme I; R, Nru I, S, Sal I. Only the Bam HI sites used to clone the genomic insert are shown. The position of alpha-2u-globulin genes are indicated by the dark boxes, and the direction of transcription is indicated by the arrows.

A


BPM
S
S
A
MA
S
B
N
FS
p2860

B


BN S M
SR
P P
A
MA B
N
FS
p2861

C

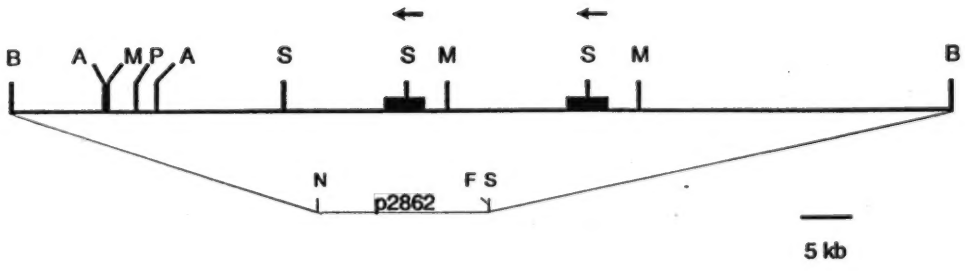
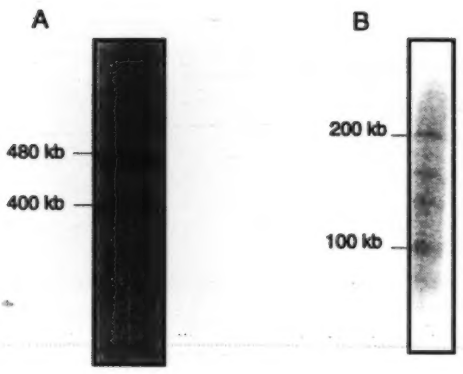

B
A
M P A
S
S M
S M
B
N
FS
p2862
5 kb

Figure 3.11 Southern hybridization analysis of rat genomic DNA separated by pulsed field gel electrophoresis.

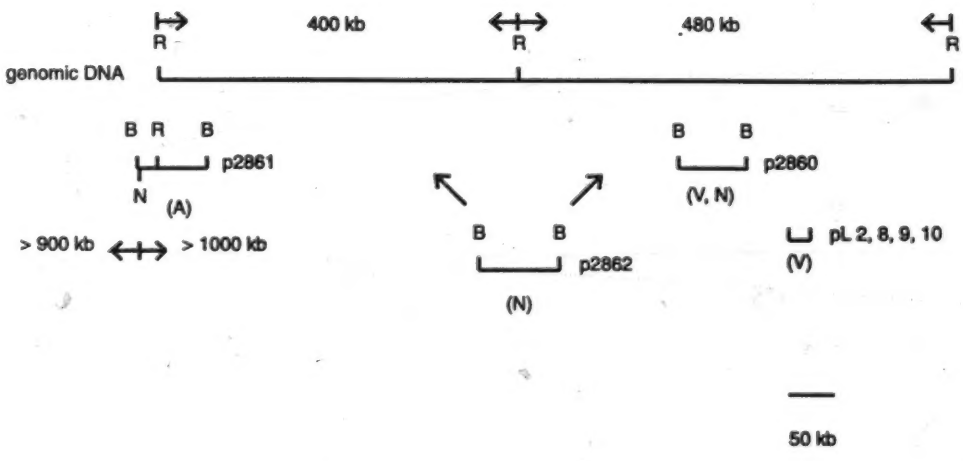
(A) Genomic DNA digested with Nru I and hybridized with the 4 kb Hind III fragment of 91R91.

(B) Genomic DNA digested with Pme I and hybridized with the same probe as in (A).

(C) Long range map of the alpha-2u-globulin locus. Restriction sites shown are: B, Bam HI; N, Not I; R, Nru I. The various genomic clones examined in this project are shown below the 880 kb region of the rat genome, defined by the 400 and 480 kb Nru I fragments, believed to contain all the alpha-2u-globulin genes. Although the exact location of p2860 and the pL 2, 8, 9, 10 series of genomic clones was not determined, it is known by virtue of their possession of class V genes (see 3.4.2) that they are on the 480 kb Nru I fragment. Genomic clone p2861 is an end clone, and has tentatively been used to anchor the end of the 400 kb Nru I fragment, since this fragment is richer in class A genes, which is the type carried by p2861. Clone p2862, because it contains only type N genes can not be assigned to a Nru I fragment. The gene classes represented by each clone are shown below each clone.


A
B
480 kb
400 kb
200 kb
100 kb

C


R
400 kb
R
480 kb
R
genomic DNA
B R B
p2861
N
(A)
> 900 kb
> 1000 kb
B B
p2860
(V, N)
B B
p2862
(N)
pL 2, 8, 9, 10
(V)
50 kb

Hybridization of the 4 kb Hind fragment from 91R91 to Southern blotted rat genomic DNA digested with Nru I produced a simple hybridization pattern (Figure 3.11). Bands of approximately 400 kb and 480 kb were apparent, suggesting all genes are present within a locus of under 880 kb, assuming the 2 Nru I fragments are adjacent in the genome. The presence of partial digestion products in the range of 900-1000 kb that hybridize to the 4 kb Hind III fragment from 91R91 tends to support the conclusion that the two Nru I fragments are adjacent in the rat genome.

The same pattern of hybridization signals were observed when probe 8 (from pL 1), and pSK2-3 were used as probes. A long range map of the alpha-2u-globulin locus prepared from the compilation of all the data obtained in this study is presented in Figure 3.11.

3.4.2. PLACEMENT OF A P1 CLONE ON THE LONG RANGE MAP

The P1 clone, p2861, can be positioned on the long range map as a consequence of a single Nru I restriction site in the genomic insert of p2861. Since all alpha-2u-globulin genes are contained on two Nru I fragments, the possible locations of the P1 clone p2861 on the long range restriction map of the cluster are limited.

Hybridization to Nru I digested and Southern blotted rat genomic DNA using the terminal 3 kb Not I fragment from p2861as a probe demonstrated hybridization to an Nru I fragment of approximately 900 kb, but not to either of the 400 or 480 kb Nru I fragments recognized by the 4 kb Hind III fragment of 91R91. Since the 3 kb Not I fragment probe hybridizes to regions outside the two Nru I fragments that define the alpha-2u-globulin locus, indicates that p2861 carries a cloned segment from one of the ends of the alpha-2u-globulin locus, as opposed to a segment that overlaps the internal Nru I site as would have been suggested by hybridization of the 3 kb Not I fragment probe to one or the other of the 400 kb or 480 kb Nru I fragments.

In an attempt to determine whether p2861 is located at the terminus of the 400 kb, or 480 kb Nru I fragment, rat genomic DNA digested with Nru I was blotted and probed using the 52 kb Nru I fragment from p2861. The resulting hybridization pattern was indistinguishable from that obtained with the 4 kb Hind III fragment from 91R91, both the 400 kb and 480 kb Nru I fragments produced a signal. A thorough dissection of p2861 will be required to identify probes specific for p2861, thus allowing for a definitive assignment of this clone to one of the alpha-2u-globulin gene containing Nru I fragments in the rat genome. Clone p2861 was tentatively assigned to the 400 kb Nru I fragment (Figure 3.11) based on the class of alpha-2u-globulin gene it carries see below (3.5.3).

3.5. ALPHA-2U-GLOBULIN GENE CLASSIFICATION

3.5.1. GENOMIC CLONE TYPES

Analysis of cloned alpha-2u-globulin sequences (Figure 2.3) indicates the presence of a single polymorphic site for each of the restriction enzymes Sst I and Vsp I in exon 4, which is within the region amplified by the PCR primers KAW1 and KAW2 (Wang, 1996). The PCR products obtained with the primer pair KAW1/KAW2 were the same size (1600 bp) when using the alpha-2u-globulin gene containing plasmid 91R91, the four alpha-2u-globulin containing lambda subclones (pL2, pL8, pL9, and pL10), and the three P1 clones (p2860, p2861, and p2862) as templates. The products could be classified into three classes, designated N, V, or A, according to their restriction endonuclease profiles for the restriction enzymes Sst I and Vsp I (Figure 3.12).

The 1600 bp product produced from plasmid 91R91, when the primer pair KAW1/KAW2 is used in the PCR, is not digested by either Sst I or Vsp I. The lack of both of these restriction sites in the amplified region of gene 91 indicates the alpha-2u-globulin gene present in this clone is a member of class N, as defined by Wang (1996).

The four lambda subclones, not surprisingly, given the fact that they represent overlapping clones of the same alpha-2u-globulin gene, all posses a Vsp I restriction site in the amplified region. Digestion with Vsp I generates restriction fragments of 1120 bp and 480 bp. The PCR products from the 4 lambda subclones did not posses a restriction site for Sst I, in keeping with the suggestion that the presence of Vsp I and Sst I restriction sites in the same gene is mutually exclusive (Wang, 1996). The presence of the Vsp I restriction site in the PCR products derived from the lambda subclones places these clones in class V.

P1 clone p2860 contains three alpha-2u-globulin genes. The restriction enzyme profile indicates the presence of genes belonging to both the N and V classes, since digestion with both Vsp I and Sst I generates a restriction profile with fragments of 480 bp, 1120 bp, and 1600 bp. The remaining P1 clones appear to each posses a single class of alpha-2u-globulin gene. The single alpha-2u-globulin gene in clone p2861 is of the A type since the PCR product from this template is cleaved by Sst I, producing restriction fragments of 1150 bp and 450 bp. The two alpha-2u-globulin genes in clone p2862 belong to class N, and lack restriction recognition sites for both Sst I and Vsp I.

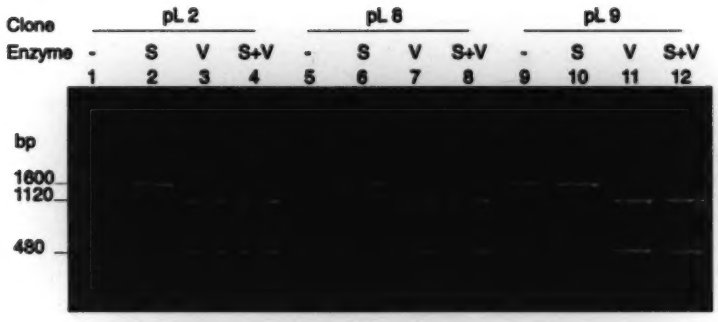
Figure 3.12 Restriction analysis of PCR products amplified from various templates with the primer pair KAW1/KAW2.

(A) KAW1/KAW2 amplified PCR products from lambda subclones pL 2 (lanes 1-4), pL 8 (lanes 5-8), and pL 9 (lanes 9-12). Samples were: undigested (-) (lanes 1, 5, 9), Sst I (S) digested (lanes 2, 6, 10), Vsp I (V) digested (lanes 3, 7, 11), or doubly digested with Sst I and Vsp I (S+V) (lanes 4, 8, 12).

(B) KAW1/KAW2 amplified PCR products from P1 clones p2860 (lanes 1-4) and p2862 (lanes 5-8). Samples were: undigested (-) (lanes 1, 5), Sst I (S) digested (lanes 2, 6), Vsp I (V) digested (lanes 3, 7), or doubly digested with Sst I and Vsp I (S+V) (lanes 4, 8).

(C) KAW1/KAW2 amplified PCR products from rat genomic DNA (lanes 1-4), P1 clone p2861 (lanes 5-8), and lambda subclone pL 10 (lanes 9-12). Samples were: undigested (-) (lanes 1, 5, 9), Sst I (S) digested (lanes 2, 6, 10), Vsp I (V) digested (lanes 3, 7, 11), or doubly digested with Sst I and Vsp I (S+V) (lanes 4, 8, 12).

A


Clone
pL 2
pL 8
pL 9
Enzyme - S V S+V - S V S+V - S V S+V
1 2 3 4 5 6 7 8 9 10 11 12
bp
1600
1120
480

B

Clone
p2860
p2862
Enzyme - S V S+V - S V S+V
1 2 3 4 5 6 7 8
bp
1600
1120
480

C

Clone
genomic DNA
p2861
pL 10
Enzyme - S V S+V - S V S+V - S V S+V
1 2 3 4 5 6 7 8 9 10 11 12
bp
1600
1150
1120
480
450

Wang (1996) conducted a similar restriction endonuclease profile analysis carried out on RTPCR products obtained from tissues of both male and female rats at various developmental stages. These results provide the opportunity to suggest possible tissues of expression for the genomic clones reported here. Table 3.3 indicates the possible expression sites for the genomic clones analyzed above.

3.5.2. PROPORTION OF EACH GENE CLASS IN THE GENOME

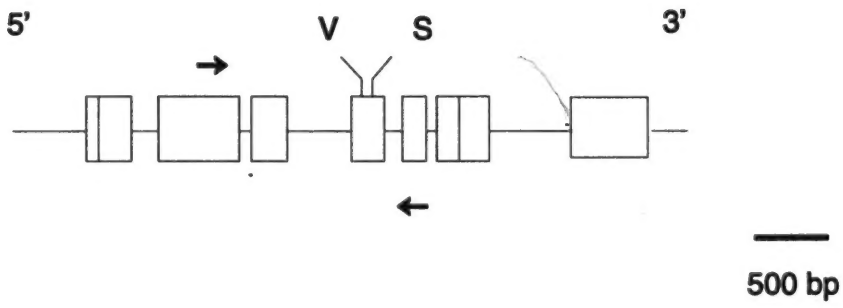
Genomic DNA prepared from male rat liver and subjected to the PCR using the primer pair of KAW1/KAW2 was also analyzed for its restriction endonuclease profile with the enzymes Sst I and Vsp I (Figure 3.12). A single PCR product of 1600 bp was obtained from genomic DNA with the primers KAW1 and KAW2. All three classes of alpha-2u-globulin genes were detected in digests of PCR products obtained from genomic DNA. Densitometric scans performed on an autoradiogram from Southern blot hybridization of digested KAW1/KAW2 amplified PCR products from genomic DNA (Figure 3.13) were used to determine an estimate for the fraction of alpha-2u-globulin genes each of the three gene classes comprise. The results of the densitometric analysis of the autoradiogram in Figure 3.13 are provided in Table 3.4. The results indicate that the majority (57%) of the alpha-2u-globulin genes belong to class N as indicated by the percent of the hybridization signal in the 1600 bp fragment following digestion with both Sst I and Vsp I. This result appears to be in keeping with the results obtained from the analysis of the three P1 clones, where at least 3 and perhaps 4 of the 6 alpha-2u-globulin genes carried by the P1 clones belong to class N. However, it seems to contrast with the frequency with which clones from class N have been isolated from cDNA and lambda genomic libraries. Of the twelve full length alpha-2u-globulin sequences deposited in Genbank, none belong to class N. This analysis suggests that approximately 30% of the alpha-2u-globulin genes belong to the A class as determined by the proportion of the hybridization signal in the 1150 bp and 450 bp fragments following digestion with Sst I. The class V genes make up about 13% of the alpha-2u-globulin genes as the 1120 bp and 480 bp fragments account for this fraction of the hybridization signal in samples digested with Vsp I.

Table 3.3. Gene Type Expression Patterns

Tissue	Gene Classes Expressed	Possible Genomic Clones Expressed
Liver	A, V, N	all
Salivary Gland	A	p2861
Preputial Gland	A, V, N	all
Kidney	A, V	pL2, 8, 9, 10 p2861 one gene on p2860
Lachrymal Gland	A, V	pL2, 8, 9, 10 p2861 one gene on p2860
Mammary Gland	A, V, N	all

Tissues examined and the classes of alpha-2u-globulin genes reported by Wang (1996) to be expressed in these tissues are reported in the first two columns of the table. The last column of the table lists potential genomic clones, from among those investigated in the present study (91R91, pL2, 8, 9, 10, and the three P1 clones p2860, p2861, and p2862), that could potentially be expressed in each tissue

Figure 3.13. Southern blot analysis PCR products. PCR products amplified from rat genomic DNA with the primer pair KAW1/KAW 2 were digested with Sst I, S (lanes 1, 4, 7); Vsp I, V (lanes 2, 5, 8); or both Sst I and Vsp I, S+V (lanes 3, 6, 9) blotted and hybridized with the PCR product obtained from 91R91 with the primer pair KAW1/KAW2. The region of the alpha-2u-globulin gene amplified by the primers KAW1/KAW2 in the PCR, the position of the Sst I and Vsp I restriction sites and the probe used in the hybridization is shown above the autoradiograph.


5'
V
S
3'
500 bp

PCR Probe

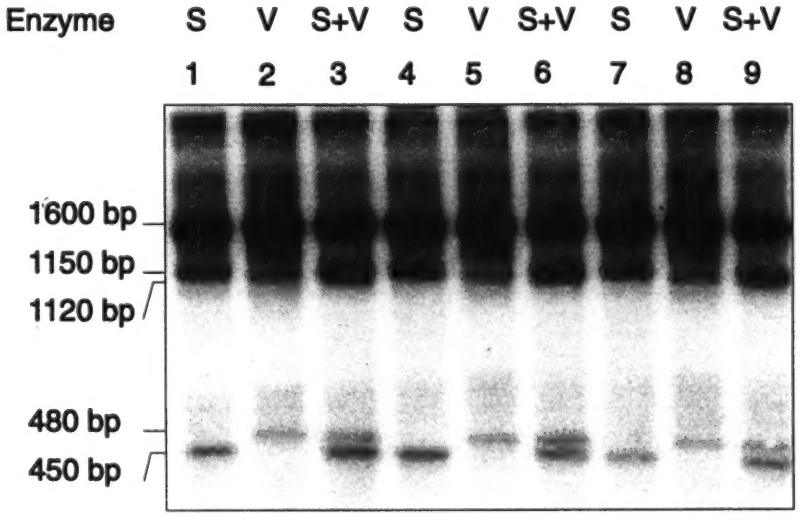

Enzyme
S V S+V S V S+V S V S+V
1 2 3 4 5 6 7 8 9
1600 bp
1150 bp
1120 bp
480 bp
450 bp

Table 3.4. Densitometric Analysis of Restriction Enzyme Digested PCR Products: Determination of Gene Class Proportions.

	Percent Signal								
Enzyme	Sst I			Vsp I			Sst I/Vsp I		
Sample	1	2	3	1	2	3	1	2	3
Fragment (bp)									
1600	69.0	68.5	73.6	83.3	85.9	83.9	54.2	55.5	61.0
1150	22.3	22.8	18.9						
							32.3	30.3	27.3
1120				12.0	10.2	11.8			
480				4.7	3.9	4.3			
							13.5	14.2	11.7
450	8.7	8.7	7.5						

The autoradiograph in Figure 3.13 was subject to a densitometric analysis. The table presents the percent of the total hybridization signal represented by each fragment in the indicated digest. The values for the 1150 bp and 1120 bp fragments, and the 480 bp and 420 bp fragments were combined for the samples that were doubly digested with Sst I and Vsp I since it was not possible to isolate the signal produced by individual fragments. The data for these fragments is provided on a line in the table between the two fragment sizes they were obtained from.

3.5.3. DISTRIBUTION OF GENE CLASSES IN THE GENOME

The distribution of the three classes of alpha-2u-globulin genes within the 880 kb alpha-2u-globulin locus was examined using Sst I and Vsp I restriction digestion of KAW1/KAW2 PCR products amplified from Nru I digested genomic DNA, isolated from slices of a pulsed field gel. The results (Figure 3.14) indicate that there is a qualitative difference between the 480 kb and 400 kb Nru I fragments with respect to their alpha-2u-globulin gene content. The results suggest that most if not all of the V class genes reside on the 480 kb Nru I fragment. Additionally, it appears that genes of the A class are distributed such that there is a higher proportion present on the 400 kb Nru I fragment than the 480 kb fragment. Taken together these two observations suggest some degree of clustering of gene types within the alpha-2u-globulin locus. An analysis of PCR products obtained from inter Nru I gel slices indicated the presence of alpha-2u-globulin sequences in the inter-band regions of the gel. This is not totally unexpected given the observation of background signal throughout the length of a lane of Nru I digested and blotted rat genomic DNA hybridized with the 4 kb Hind III fragment from 91R91. This likely represents the end-result of unavoidable DNA shearing.

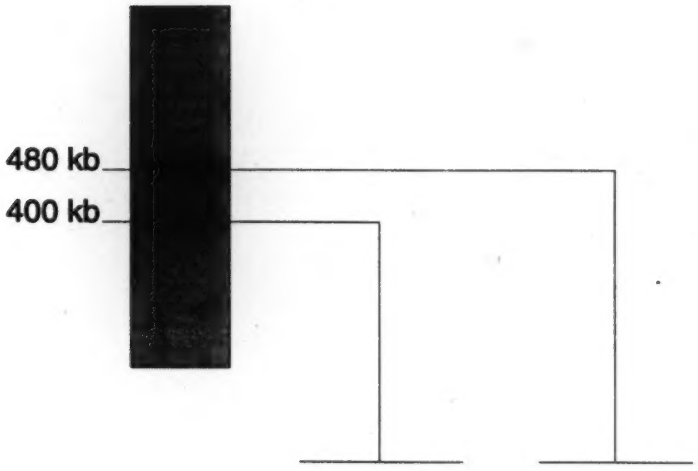
Given the distribution of gene classes over the two Nru I fragments it is quite likely that the alpha-2u-globulin gene represented by the lambda subclones pL 2, 8, 9, 10 resides on the 480 kb Nru I fragment.

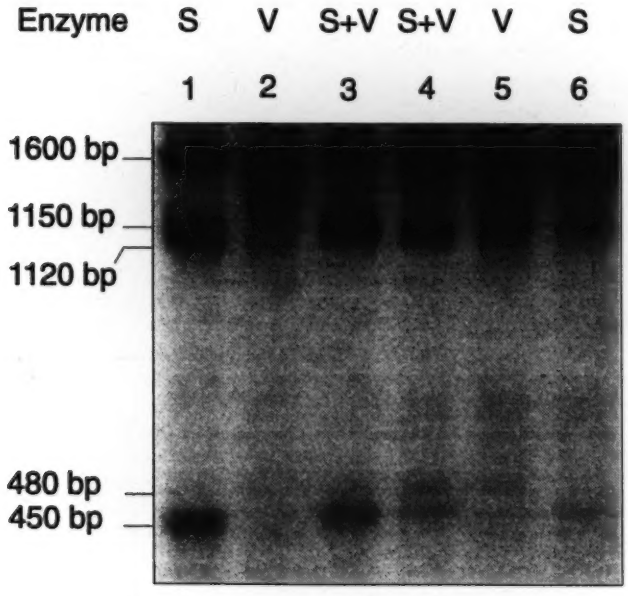
The P1 clone p2860, again because one of the genes carried by this clone belongs to class V, can tentatively be placed on the 480 kb Nru fragment. Because p2860 does not posses restriction sites for Nru I it is unlikely to overlap both the 480 kb and 400 Nru I fragments.

Clone p2861, as a consequence of the presence of a restriction site for Nru I, either represents a cloned segment from one of the termini of the gene cluster, or overlaps the two Nru I fragments. The results of Southern hybridization using the 3 kb Not I fragment of p2861 as a probe suggested that the insert in p2861 is derived from one of the ends of the alpha-2u-globulin locus (3.4.2).

The gene carried by p2861 is an A type gene. Genes belonging to class A are located on both the 400 kb and 480 kb Nru I fragments, however, the distribution is disproportionate, with more A class genes located on the 400 kb Nru I fragment. A definitive assignment to one of the two Nru I fragments is not possible based on gene type, but given the disproportionate distribution of the class A genes over the two fragments, it could be argued that the genomic insert in p2861 may represent the terminal region of the 400

Figure 3.14. Southern analysis of PCR products from pulsed field gel purified templates. The top portion of the figure shows a lane of a pulsed field gel containing Nru I digested DNA, that was subject to Southern hybridization with the 4 kb Hind III fragment from 91R91. Using this autoradiogram as a guide, Nru I digested DNA from the adjacent lane was excised from the gel at positions corresponding to the hybridization signals from the 400 kb (lanes 1-3) and 480 kb (lanes 4-6) Nru I fragments. DNA from the excised gel slices was amplified in a PCR reaction with KAW1/KAW2, digested with Sst I, S (lanes 1, 6); Vsp I, V (lanes 2, 5); or both Sst I and Vsp I, S+V (lanes 3, 4) and hybridized with the PCR product obtained from 91R91 with KAW1/KAW2.


480 kb
400 kb


Enzyme S V S+V S+V V S
1 2 3 4 5 6
1600 bp
1150 bp
1120 bp
480 bp
450 bp

kb Nru I fragment, although the exact location of p2861 has not been established.

The two genes present on p2862 are both of the N type, as is gene 91 carried by 91R91. Given the observation that genes belonging to class N are localized to both the 400 kb and 480 kb Nru I fragments, it is not possible to assign p2862, or gene 91 to either of the Nru I fragments.

A summary of the possible position of each genomic clone in the alpha-2u-globulin cluster is presented in Figure 3.11.

CHAPTER 4.
RESULTS
CYTOGENETIC ANALYSES

4.1. RAT METAPHASE FISH

4.1.1. LOCALIZATION OF THE ALPHA-2U-GLOBULIN LOCUS

Fluorescent *in situ* hybridization to metaphase spreads and interphase nuclei prepared from Rat 2 fibroblasts, using biotinylated 91R91 as a probe, resulted in the generation of a hybridization signal pattern consistent with a single chromosomal locus for the alpha-2u-globulin genes (Figure 4.1). This is consistent with linkage analysis (Nikaido et al., 1982), somatic cell hybrid analysis (Szpirer et al., 1990), and *in situ* hybridization with a radiolabeled probe (Kurtz, 1981) all of which have indicated that the genes for the alpha-2u-globulins were clustered on rat chromosome 5. The observed hybridization signal was localized to the long arm of chromosome 5, and more specifically to the chromosomal region 5q22→5q24 as determined by the Q-like-banding pattern evident as a result of staining with DAPI.

4.2. P1 FIBER FISH

4.2.1. P1 GENE CONTENT

Two color Fiber FISH to extended P1 DNA targets generated direct visual evidence demonstrating the number of alpha-2u-globulin genes on each of the P1 clones. The alpha-2u-globulin gene content of each of the P1 clones was determined following hybridizing the linearized P1 target DNAs simultaneously with digoxigenin labeled 91R91 and the respective P1 DNA labeled with biotin. Following the detection scheme outlined in section 2.5.6, alpha-2u-globulin genes were visualized as red signals at discrete locations along the linearized P1, which was "counter-stained" as a result of hybridization with the biotinylated P1 DNA, and visualized in green. Clone p2860 demonstrated the presence of three genes, p2861 a single gene, and p2862 two alpha-2u-globulin genes (Figure 4.2). These results confirmed the results for gene number obtained by Southern hybridization (see 3.3.3).

Figure 4.1 Rat metaphase FISH.

(A) Rat metaphase spread stained with DAPI demonstrating a Q-like banding pattern. Chromosome 5 is indicated by the arrows.

(B) Same spread as in (A) showing the hybridization of biotinylated 91R91 at 5q22→24 as a yellow signal on propidium iodide counterstained chromosomes.

(C) ideogram of rat chromosome 5 indicating the location of the alpha-2u-globulin gene family.

C

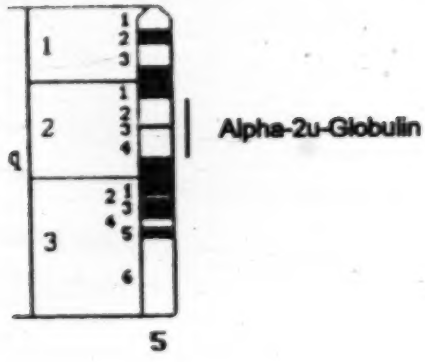


Figure 4.2. P1 Fiber FISH.

(A) FISH to a p2860 DNA fiber with biotinylated p2860 visualized in green, and digoxigenin labeled 91R91 visualized in red. The position of the genes is indicated by the arrows

(B) FISH to a p2861 DNA fiber with biotinylated p2861 visualized in green, and digoxigenin labeled 91R91 visualized in red. The position of the gene is indicated by the arrow.

(C) FISH to a p2862 DNA fiber with biotinylated p2862 visualized in green, and digoxigenin labeled 91R91 visualized in red. The position of the genes is indicated by the arrows.

A 1μm bar is shown in the bottom right of each panel. The bar represents approximately 3.8 kb in panels (A) and (B) and approximately 4.2 kb in panel (C).

p2860
A

p2861
B

p2862
C

4.2.2. SPACING OF ALPHA-2U-GLOBULIN GENES CARRIED ON P1 CLONES

Fiber FISH, in addition to directly demonstrating gene content, allows for the semi-quantitative analysis of both the distance between multiple genes on the same P1 molecule, and their distance from the ends of the linearized target DNA. The data used to compile the Fiber FISH maps for the three P1 clones are provided in the appendix.

Following FISH to linearized p2860, nine molecules were selected and analyzed to determine the physical arrangement of alpha-2u-globulin genes in this clone. For the nine spreads of p2860 from which measurements were obtained, the average degree to which the DNA was stretched was 0.266 μm/kb which is approximately 80% of that expected for relaxed DNA (0.34 μm/kb). The degree of DNA stretching ranged from 0.152-0.389 μm/kb. The apparent location of each alpha-2u-globulin gene on each of the nine molecules examined is shown schematically in Figure 4.3.

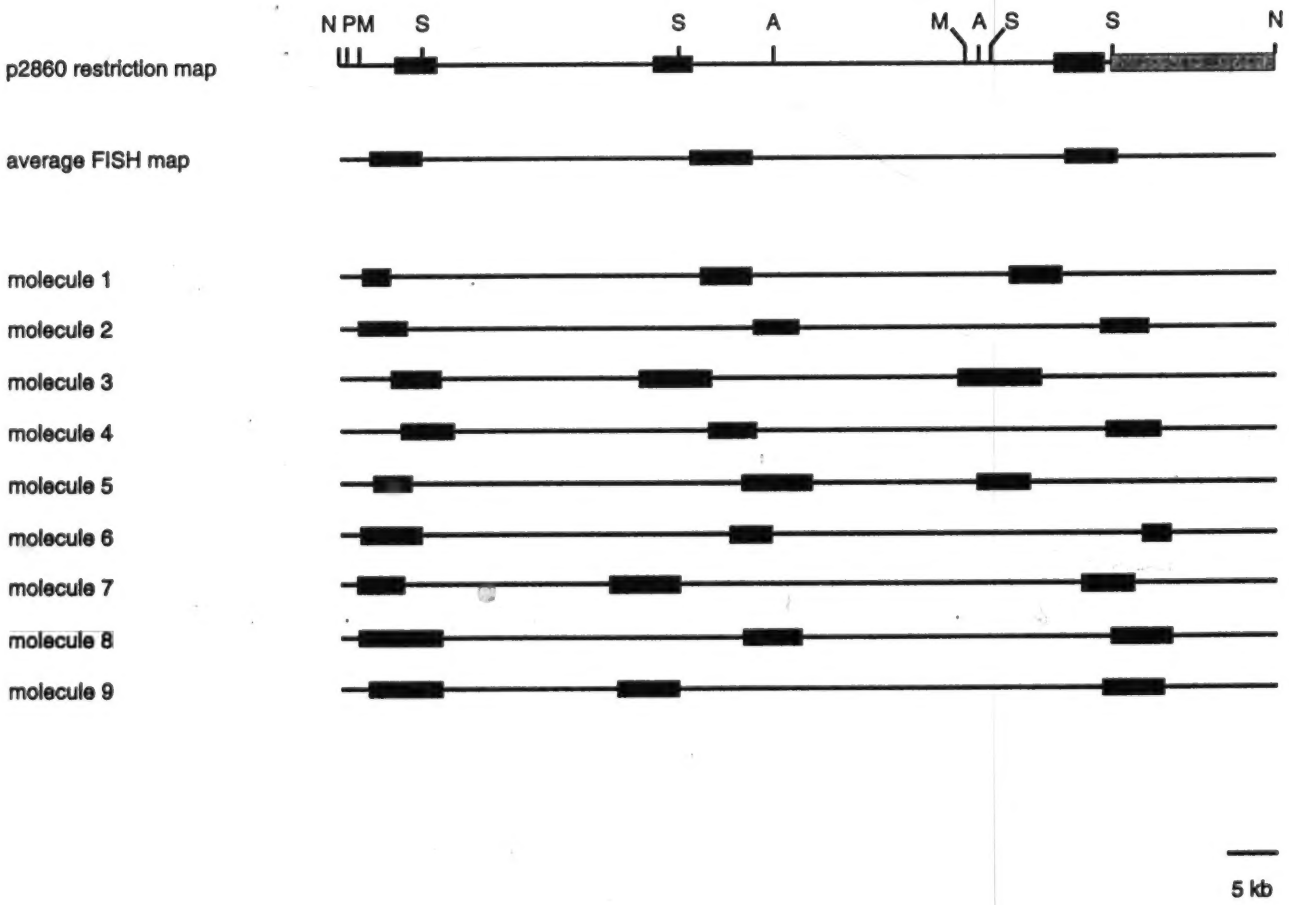
The genes adjacent to the termini of the linearized clone have been positioned approximately 15.2 kb from the end containing vector sequences following linearization with Not I, and approximately 2.9 kb from the other end. These compare to values of approximately 20 kb and 5.5 kb respectively, as determined by restriction endonuclease and Southern hybridization analyses (Figure 3.9).

The intergene distances were also calculated from the Fiber FISH results. The distance between the alpha-2u-globulin gene adjacent to the vector sequences in the linearized clone, and the next gene along the P1 was estimated to be 30.6 kb, while the distance between the second gene and the gene located at the other terminus was determined to be about 26.3 kb. These values are comparable to the estimates of 32.5 kb and 21 kb for the same two intervals based on restriction enzyme digestion and Southern hybridization (Figure 3.9).

Ten molecules were examined in the analysis of clone p2861. The average degree of stretching (0.266 μm/kb) observed for these 10 molecules was similar to that reported above for p2860, and was approximately 80% of that expected from relaxed DNA. The degree of DNA stretching ranged from 0.188-0.350 μm/kb.

The single alpha-2u-globulin gene on p2861 was found to map approximately 64 kb in from the end containing vector sequences on the linearized clone, and approximately 25 kb in from the other end. Based on these values, which delimit the position of the ends of the gene with respect

Figure 4.3. Schematic representation of fiber FISH to nine p2860 molecules. The top line of the figure shows the restriction map of p2860 as indicated in Figure 3.9. Restriction sites shown are: A, Sst II; M, Sma I; N, Not I; P, Pme I; S, Sal I. Vector sequences are shaded in grey, and alpha-2u-globulin genes are represented by the dark boxes. The second line of the figure represents the average Fiber FISH map derived from the nine molecules drawn below.


N P M S S A M A S S N
p2860 restriction map
average FISH map
molecule 1
molecule 2
molecule 3
molecule 4
molecule 5
molecule 6
molecule 7
molecule 8
molecule 9
5 kb

to the ends of the P1 clone, and the size of the P1 clone (94 kb), the size of the sequences detected by the probe is estimated to be 5.4 kb, which is in good agreement with the known size of the insert in 91R91 (4.5 kb). The positioning of the gene using Fiber FISH is also in good agreement with the localization of the gene by Southern hybridization (Figure 3.9). The ten P1 molecules considered in the physical analysis of p2861 are depicted in Figure 4.4.

A total of eleven molecules were examined in the analysis of p2862 by Fiber FISH. The two genes on p2862 are predicted to be separated by about 18.8 kb, which is comparable to the 13 kb determined for the same interval on the basis of restriction enzyme mapping and Southern hybridization. The physical distances between the genes and the ends of the linearized clone were also determined from the Fiber FISH images. The gene adjacent to vector sequences following linearization with Not I was found to be 45.5 kb from the end, while the alpha-2u-globulin next to the other terminus was approximately 27.7 kb in from the end. The values determined from the Fiber FISH images of p2862 molecules correlates well with the interval sizes determined by restriction mapping. The eleven molecules of p2862 considered in the analysis are diagrammed in Figure 4.5. The average degree of DNA stretching observed for these 11 molecules was 0.237 μm/kb, with a range of 0.178-0.303 μm/kb.

4.3. RAT GENOMIC DNA FIBER FISH

4.3.1. ALPHA-2U-GLOBULIN GENE ORIENTATION

Although the molecular and cytogenetic analysis of the three P1 clones increases our understanding of the organization of the alpha-2u-globulin locus, the conclusions that can be made are only applicable to the small segments of the locus covered by these clones. In an attempt to extend our understanding of the organization to the entire alpha-2u-globulin locus the Fiber FISH methodology used to examine the P1 clones was applied to rat genomic DNA.

A total of 50 independent signal profiles, obtained from Fiber FISH images of rat genomic DNA, hybridized with biotinylated 91R91, to identify the 3′ end of alpha-2u-globulin genes, and digoxigenin labeled pSK2-3, to identify the 5′ end of alpha-2u-globulin genes, were analyzed (Table 4.1). The profiles indicate the orientation (head to tail, head to head, etc.) of groups of alpha-2u-globulin genes, observed in the same region of a field of view under the microscope, according to the observed hybridization pattern.

Figure 4.4. Schematic representation of fiber FISH to ten p2861 molecules. The top line of the figure shows the restriction map of p2861 as indicated in Figure 3.9. Restriction sites shown are: A, Sst II; F, Sfi I; M, Sma I; N, Not I; P, Pme I; R, Nru I; S, Sal I. Vector sequences are shaded in grey, and alpha-2u-globulin genes are represented by the dark boxes. The second line of the figure represents the average Fiber FISH map derived from the nine molecules drawn below.

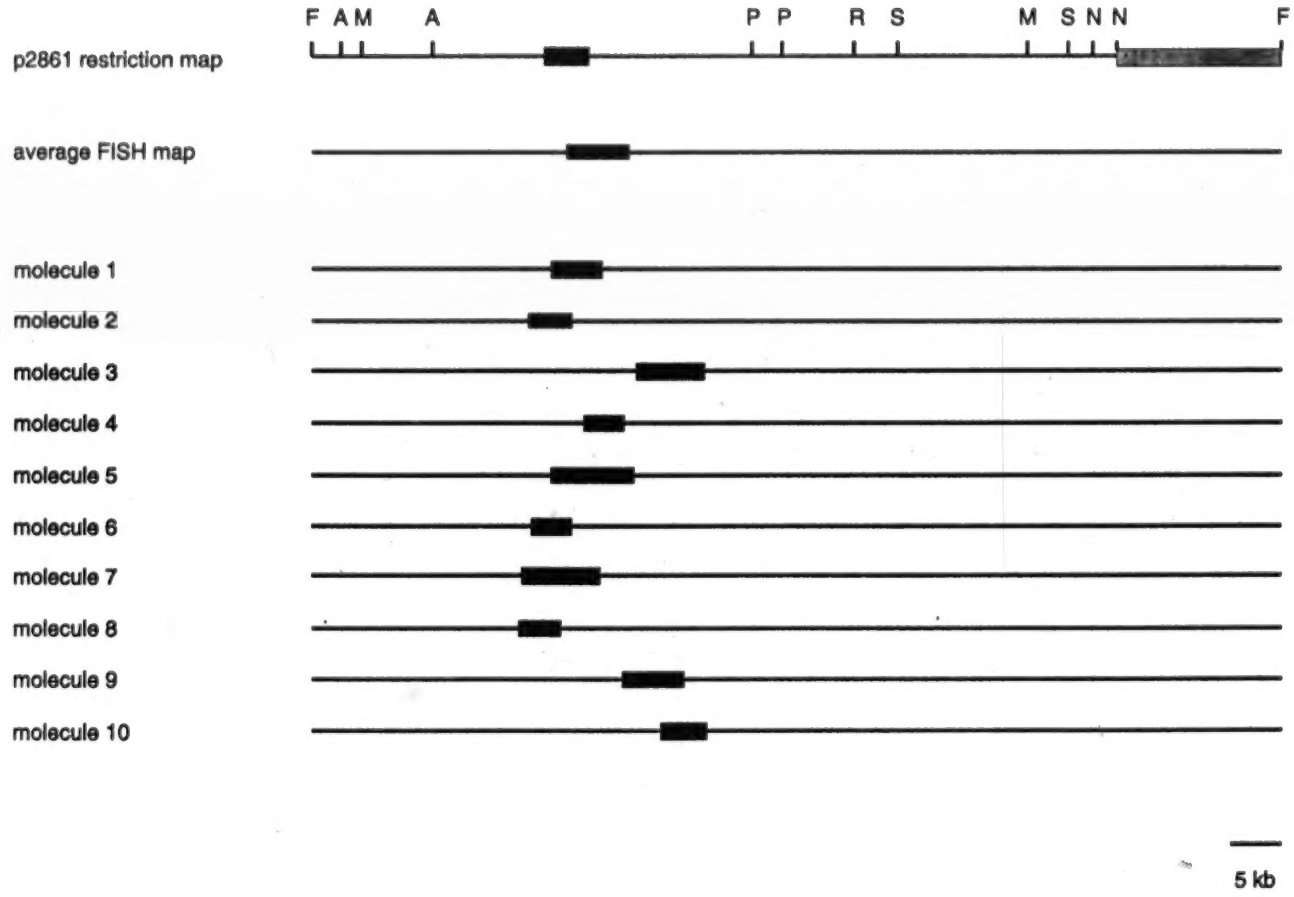

F A M A P P R S M S N N F
p2861 restriction map
average FISH map
molecule 1
molecule 2
molecule 3
molecule 4
molecule 5
molecule 6
molecule 7
molecule 8
molecule 9
molecule 10
5 kb

Figure 4.5. Schematic representation of fiber FISH to eleven p2862 molecules. The top line of the figure shows the restriction map of p2862 as indicated in Figure 3.9. Restriction sites shown are: A, Sst II; M, Sma I; N, Not I; P, Pme I; S, Sal I. Vector sequences are shaded in grey, and alpha-2u-globulin genes are represented by the dark boxes. The second line of the figure represents the average Fiber FISH map derived from the nine molecules drawn below.

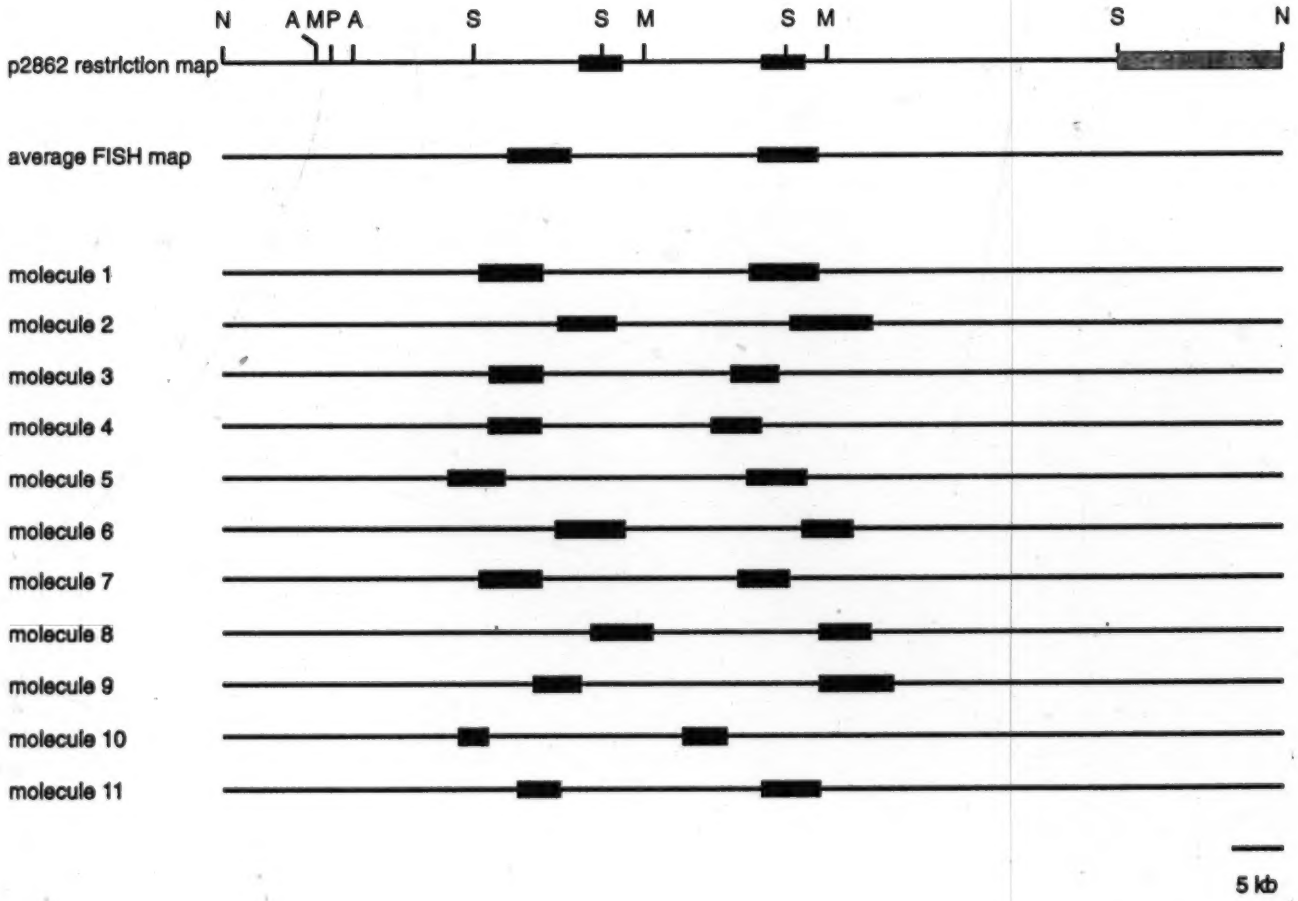

N
A
M
P
A
S
S
M
S
M
S
N
p2862 restriction map
average FISH map
molecule 1
molecule 2
molecule 3
molecule 4
molecule 5
molecule 6
molecule 7
molecule 8
molecule 9
molecule 10
molecule 11
5 kb

Table 4.1. Alpha-2u-Globulin Gene Orientation in the Rat Genome

Profile	Gene Orientation
1	HT HT HT HT HT HT HT
2	HT HT HT HT HT
3	HT HT HT HT
4	HT HT HT TH
5	HT HT HT TH
6	HT HT HT HT TH HT
7	HT HT HT HT TH HT
8	HT HT HT HT HT HT HT TH HT TH TH
9	HT HT HT TH
10	HT HT HT HT TH
11	TH HT HT HT HT T TH HT HT
12	TH HT HT HT HT T TH HT
13	HT T TH HT HT
14	HT T TH HT HT
15	TH TH HT HT HT HT
16	HT TH HT TH TH HT
17	TH HT HT HT HT
18	HT TH TH HT HT HT
19	HT TH TH HT HT
20	HT HT HT TH
21	TH TH HT HT
22	HT HT HT
23	HT HT HT HT
24	HT HT TH
25	HT HT HT HT
26	HT HT TH HT
27	HT HT HT
28	HT HT HT HT T TH HT HT HT TH
29	TH TH HT HT
30	HT HT HT HT HT
31	T TH HT HT
32	HT HT HT HT
33	HT HT HT HT
34	TH TH HT
35	TH TH HT
36	HT HT HT HT HT
37	TH TH HT
38	TH HT HT
39	HT HT HT
40	HT HT HT HT TH
41	HT HT TH
42	HT TH TH
43	HT TH HT
44	HT HT HT TH
45	TH HT TH
46	TH HT HT HT
47	HT HT TH HT TH TH
48	HT HT HT HT HT
49	TH HT TH TH
50	HT HT HT TH

The observed orientation of alpha-2u-globulin genes in 50 independent Fiber FISH profiles: H (head, 5´ end), T (tail, 3´ end).

The 50 profiles were assembled into a large gene array representative of the alpha-2u-globulin locus (Figure 4.6). This was achieved by taking several of the longer profiles and aligning them according to overlapping gene orientation patterns. The relatively low degree of complexity in the orientation patterns allowed for the construction of the array by hand. Shorter profiles were then added to the backbone prepared from the longer profiles at appropriate positions. It should be noted that many of the smaller profiles could be placed at several positions in the array, although only a single position is indicated. There is the possibility that the array may extend further in either direction. This possibility is especially evident at the left end of the array where the genes are all tandemly repeated.

The organization for the cluster presented in Figure 4.6 represents one possible arrangement for the genes in the alpha-2u-globulin cluster, and is not a definitive assessment of the structure of the locus. More confidence can be placed in the gene orientations that are repeatedly observed among the shorter profile segments than the assembled larger array. The size of the array, as determined by the 50 profiles reported here is 20 genes, which accounts for the majority of the genes based on the size estimates of the gene cluster made by solution hybridization kinetics (Kurtz, 1981) and slot blot analysis (this work). Six sample rat genomic Fiber FISH profiles that, taken together account for the entire alpha-2u-globulin array proposed are shown in Figure 4.7.

In the proposed arrangement the vast majority of the genes are present in tandem, however, there are examples of genes demonstrating head to head linkage, and tail to tail linkage distributed at several locations throughout the cluster, such that the array is not entirely homogenous with respect to gene orientation.

Figure 4.6. A possible arrangement of the genes at the alpha-2u-globulin locus. The 50 independent gene orientation profiles reported in Table 4.1 were assembled into an array according to overlapping hybridization signal patterns. Symbols are: H, head (5′) end of the gene; T, tail (3′) end of the gene. For one gene in the array it was not possible to determine the orientation, it is designated with a "T".

| Profile | | |

Figure 4.7. Six rat genomic Fiber FISH profiles covering the alpha-2u-globulin locus.

(A) Fiber FISH profile 36 (HT-HT-HT-HT-HT)

(B) Fiber FISH profile 20 (HT-HT-HT-TH)

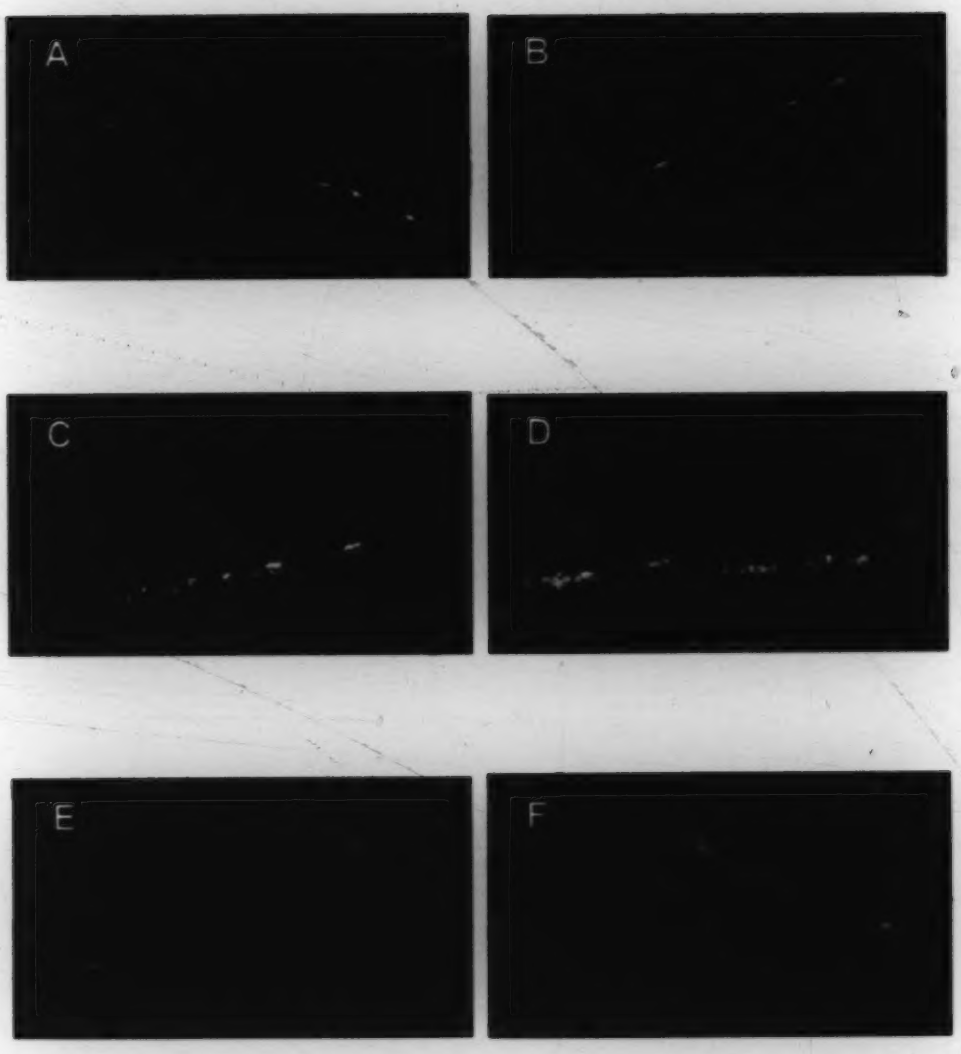
(C) Fiber FISH profile 47 (HT-HT-TH-HT-TH-TH)

(D) Fiber FISH profile 18 (HT-TH-TH-HT-HT-HT)

(E) Fiber FISH profile 13 (TH-TH-HT-T-TH)

(F) Fiber FISH profile 41 (HT-HT-TH)

In all panels the 5′ end of the gene (H) is visualized by a red signal, and the 3′ end of the gene (T) is visualized by a green/yellow signal.


A
B
C
D
E
F

CHAPTER 5. DISCUSSION

5.1. LOCATION OF ALPHA-2U-GLOBULIN AND OTHER LIPOCALIN GENES

This work refines the position of the alpha-2u-globulin gene cluster to chromosomal region 5q22→24 by FISH. The rat alpha-2u-globulin locus had previously been assigned to chromosome 5q by somatic cell hybrid panel screening and *in situ* hybridization (Kurtz, 1981; Szpirer et al., 1990). Furthermore, most, if not all of the alpha-2u-globulin genes were found to be contained on Nru I restriction fragments of about 400 kb and 480 kb. This result indicates that the entire cluster is contained within a 900 kb region of the rat genome.

Whether or not this cytogenetic region of the rat genome contains other lipocalin genes is not clear at present. The only other lipocalin gene thus far assigned to rat chromosome 5 is that of alpha-1-acid glycoprotein. In the mouse, both alpha-1-acid glycoprotein and alpha-1-microglobulin have been mapped 2.8 cM distal to MUP on chromosome 4. The gene for alpha-1-microglobulin has not been mapped in the rat, so the possibility remains that a similar cluster consisting of these three lipocalins may exist on rat chromosome 5. Indeed this is quite likely, for two reasons. First, there are numerous instances of conserved syntenies beween mouse chromosome 4 and rat chromosome 5 (Levan et al., 1991; MGD). Second, mouse chromosome 4 shares cytogenetic band-by-band homology with the distal 90% of rat chromosome 5 (Nesbitt, 1974; Yoshida, 1978; Szpirer et al., 1988; Levan et al., 1991) indicating the potential for genetic homology in the region, as has been reported for other regions demonstrating cytogenetic banding pattern homology in the rat and mouse (Levan et al., 1991).

Four lipocalin family members have been mapped to mouse chromosome 2 which shares morphological identity, and gene homologies with rat chromosome 3. The rat homologues of the lipocalin genes that map to mouse chromosome 2 have not been mapped, however, several other genes that flank the cluster of four genes in the mouse have been mapped to rat chromosome 3 (RATMAP), suggesting that genes found between these flanking markers in the mouse will also map to rat chromosome 3. It will be interesting to see, as gene mapping progresses in the rat, if this prediction proves correct.

It is likely that the rather large cluster of seven lipocalin genes at 9q34 in humans, more closely represents the arrangement in the common mammalian ancestor, and that ancestral linkage map has been broken up to a larger extent in the rodent lineages. Evidence of this type of fragmentation has been reported as a common feature of rodent genomes (Oakey et al., 1991; O'Brien, and Graves, 1991; Lalley et al., 1991).

5.2. HOMOLOGY AMONG ALPHA-2U-GLOBULIN GENES

5.2.1. SEQUENCE SIMILARITY

The sequences for the alpha-2u-globulin family members reported in Figure 2.3 display a high degree of sequence identity. The results from genomic Southern hybridization, and genomic library screening (3.1.1 and 3.2.2) confirm and extend the conclusions made from the limited sequence data available. This is a common feature of many multigene families (Arnheim, 1983). It not only suggests that these genes arose by gene duplication events, but that either this amplification has occurred in recent evolutionary time, or the genes are being acted upon in a fashion that reduces variation between family members (see below).

Among the 12 sequences retrieved from Genbank, nucleotide identities among the exonic regions ranged from 77% (exon 3) to 97% (exon 5). Similarily high degrees of sequence identity have been noted for alpha-2u-globulin intronic sequences, ranging between 90%-93% (Yamamoto et al., 1989), and 5' flanking sequences, approximately 90% identity (Yamamoto et al., 1989; Winderickx et al., 1987; Wang, 1996).

It is intriguing that exons 6 and 7, display such a high degree of conservation among the twelve gene sequences reported here, considering these are untranslated regions of the alpha-2u-globulin gene. This may indicate a potential regulatory function for these sequences. Surprisingly, exon 3 possesses the lowest degree of sequence identity among members of this gene family. This may indicate that exon 3 codes for a protein domain that imparts a tissue specific function. This is supported, in a limited manner, by similarity between sequences expressed in the same tissue. Comparison between available alpha-2u-globulin sequences (Figure 2.3) indicates that the two genes known to be expressed in the submaxillary gland share nucleotide variants in exon 3 not shared by the hepatically expressed genes. The decreased sequence identity between family members in exon 3 may permit the derivation of oligonucleotide probes useful for predicting the tissue of expression of the given gene.

A total of 87 positions demonstrated nucleotide variation among the 12 sequences compared in figure 2.3. Of these 87 variant positions, 50 shared the variant nucleotide with at least one of the other twelve sequences compared, indicating that recombination events between family members are a common event. The fraction of variant nucleotide positions that share identity with the nucleotide at the same position in at least one other gene ranges from 100% in exon 5 to a low of 41% in the seventh exon. In general, the fraction of variant positions sharing identity with the nucleotide at the homologous position in another gene is 15-20% lower in the sixth and seventh exons, which are non-coding, compared to the other exonic regions.

Evidence of divergence between the 3′ ends of different alpha-2u-globulin genes was observed in the present study (3.2.4). Results of genomic Southern hybridization using probes derived from adjacent regions in the 3′ end of gene 91 indicated that there is some variation in this region. This variation may serve a regulatory function, for example, by altering mRNA stability or translation to aid in establishing the diverse expression patterns observed for the various members of the gene family.

The similar structure of the alpha-2u-globulin genes coupled with their close physical linkage may facilitate sequence exchanges, via gene conversions, between family members. Such events may favor the exchange between specific gene sets with similar patterns of expression. These same features combined with the diverse expression patterns observed for different family members make this an important system for the study of the molecular mechanisms of regulatory evolution.

5.2.2. CROSS-HYBRIDIZATION OF FLANKING SEQUENCES AMONG ALPHA-2U-GLOBULIN GENES

Southern hybridization to rat genomic DNA, in addition to restriction endonuclease mapping, has indicated a high degree of conservation between upstream regions of alpha-2u-globulin genes. This is supported by results obtained from genomic library screens performed with probes derived from the upstream region of gene 91 indicating that homology between alpha-2u-globulin genes may extend up to 11-12 kb. A similar conservation of sequence was observed in the 3' flanking regions of alpha-2u-globulin genes.

The extensive homology observed between the flanking sequences of alpha-2u-globulin genes is not an uncommon feature of multigene families. The members of the human U1 RNA gene family, for example, share a high degree of sequence homology at distances of over 24 kb upstream of the gene and 20 kb downstream of the gene (Bernstein et al., 1985).

It was also observed that a probe derived from sequences between +900 and +4100 bp of gene 91 weakly hybridized to mouse genomic DNA. This may be fortuitous hybridization, or it may indicate that alpha-2u-globulin and mouse MUP genes also share a certain degree of homology in their upstream regions. The later possibility is favored since Winderickx et al. (1987) demonstrated, by direct sequence comparison, that the alpha-2u-globulins and the MUPs share a high degree of homology in their promoter regions. The above results suggest that this homology may extend beyond the gene proximal promoter sequences.

The relatively few differences in nucleotide sequences in the upstream regions of the members that have been characterized would suggest that the developmental, hormonal and tissue-specific expression patterns observed for various subsets of alpha-2u-globulin genes result from a small number of simple nucleotide substitutions, or alternatively that the expression of subsets of similarily expressed alpha-2u-globulin genes are regulated by a locus control region.

The finding of a large duplicated region in inverted orientation flanking two alpha-2u-globulin genes has interesting implications for the organization of the locus. Intrachromosomal recombination between homologous elements in opposite orientation results in the inversion of the DNA segment between the oppositely oriented elements. If these inverted segments commonly flank alpha-2u-globulin genes, as might be expected in a homogenous multigene family, such as the alpha-2u-globulins, then they could certainly be involved in generating the localized gene inversions noted in the alpha-2u-globulin locus, which is predominantly organized as direct tandem repeats.

These regions of inverted symmetry may have also been involved in the ancestral inversion event that generated the 45 kb MUP gene pallindrome. An approximately 4.5 kb region in inverted orientation is found at each end of the 45 kb MUP unit in the mouse. The possibility exists that these elements which have been termed C1 and C2 (see Figure 1.4) (Bishop et al., 1985) are evolutionary remnants of the present day inverted repeats that flank the alpha-2u-globulin genes.

5.3. ORGANIZATION OF THE ALPHA-2U-GLOBULIN GENE CLUSTER

Analysis of the P1 clones reported in this study, coupled with the results of Fiber FISH to rat genomic DNA indicated that the majority of the alpha-2u-globulin genes are organized as tandem head to tail repeats. This is in contrast to the expectation that the organization of the alpha-2u-globulin gene cluster would resemble that of the mouse MUP locus.

Analysis of the two P1 clones, p2860 and p2861, which each contained multiple alpha-2u-globulin genes indicated that the spacing between alpha-2u-globulin genes was variable. Intergene distances ranged from about 13 kb to over 30 kb.

The repeat units of a tandem array are often nearly identical as evidenced by the U1 RNA and U2 RNA genes in sea urchins (Card et al., 1982). However, polymorphisms in the spacer regions between genes ranging from point mutations to large insertions and deletions are also common among tandemly repeated genes (Bernstein et al., 1985; Hentschel and Birnsteil, 1981). It remains to be determined whether the flanking homology surrounding the alpha-2u-globulin gene extends to the borders of the adjacent genes, or whether the flanking homologies gradually deteriorate on either side of the gene such that individual alpha-2u-globulin genes exist as islands of conserved sequence clustered in an 880 kb region of unique DNA.

Evidence from gene families like the globins, keratins, and the homeotic genes suggest that there can be functional significance to the chromosomal arrangement of genes. During development, the globin genes are expressed in the same order in which they are linearly arrayed along the chromosome (reviewed by Evans et al., 1990), the order of expression of three keratin genes in stratified epithelia is also the same as their order along the chromosome (Filion et al., 1994), and the order of the homeotic genes reflects their spatial domains of expression (Krumlauf, 1992). Since the various alpha-2u-globulin genes are expressed in a tissue, developmental, and hormonally regulated fashion, it is of interest to determine whether the alpha-2u-globulins possess a genomic organization that is at all related to the regulatory programs they are subject to.

We have been unable to identify distinct classes of genes using differential hybridization of genomic Southern blots, and genomic libraries. A similar protocol had revealed two major classes of MUP genes: the functional group 1 genes, and the group 2 pseudogenes. Kurtz (1981) reported the isolation of two genomic clones with 3' ends that were divergent enough to prevent hybridization with that represented by gene 91. The prevalence of genes with 3' ends divergent from that of gene 91 was examined by screening a genomic library with probes derived from two adjacent regions of the 3' end of gene 91, one of which was the same as the probe which failed to hybridize to two of the genomic clones isolated by Kurtz (1981). Results indicated that there are a few, likely less than 5, alpha-2u-globulin genes with 3' ends divergent from gene 91. The divergence appears to be in regions of the gene that are noncoding. Whether the genes with divergent 3' ends represent a class of genes that are in close proximity within the cluster, or posses a distinct regulatory pattern is unclear.

Classes of alpha-2u-globulin genes have been detected on the basis of restriction site polymorphisms (Gubits et al., 1984; Wang, 1996). Evaluation of the genomic clones utilized in this study, as well as the rat genome as a whole, was performed using the diagnostic restriction enzymes Sst I and Vsp I. The proportion of the three classes of genes that can be identified in this fashion were determined.

Class N, which lack both restriction sites accounts for 57% of the genes in the genome, but none of the 12 alpha-2u-globulin sequences retrieved from a Genbank search (Figure 2.3). Gene 91, plus at least one of the genes carried by p2860, and both genes on p2862 belong to class N. Therefore, unlike the under-representation in Genbank, class N genes account for the majority of genes examined in this study.

Genes from class N show the most limited tissue expression pattern, being found only in the liver, preputial gland and mammary gland. However in the preputial gland, which synthesizes the most alpha-2u-globulin, Class N genes account for the majority of genes expressed. Class N genes also represent a significant proportion of the genes expressed in the liver, which synthesizes milligram quantities of alpha-2u-globulin daily (Roy, 1979). It appears then that expression of class N genes is associated with tissues synthesizing substantial quantities of alpha-2u-globulin.

Approximately 30% of the alpha-2u-globulin genes were classified as type A genes. Class A genes were expressed to some extent in all tissues that express alpha-2u-globulin, and were the only type expressed in the salivary glands (Wang, 1996). Of the sequences retrieved from Genbank, 3 of the 12 were A type genes. The gene carried by p2861 also belongs to class A.

The remaining 13% of the alpha-2u-globulin genes are accounted for by class V. Class V genes are expressed in all alpha-2u-globulin expressing tissues except the salivary gland. Nine of the sequences reported in Genbank are class V genes, as is one of the genes on p2860, and the gene represented by clones pL 2, 8, 9, 10.

The three classes of genes are present in the genome in an approximately 4:2:1 (N:A:V) ratio. One possibility suggested by this observation is that a cluster of 4 N class genes, 2 A class genes, and a V class gene served as a template in the expansion to the present day alpha-2u-globulin locus. Analysis of the 400 and 480 kb Nru I fragments, which contain all of the alpha-2u-globulin genes, for the classes of genes located on them, however, suggested that there is some degree of clustering of gene types in the genome. All class V genes are located on the 480 kb Nru I fragment, and there is an unequal distribution of A type genes, with most located on the 400 kb Nru I fragment. Genes belonging to class N are

distributed approximately equally between both the Nru I fragments. The observed gene class distribution is not consistent with the generation of the alpha-2u-globulin locus from the expansion of an ancestral cluster containing N class, A class, and V class genes in a 4:2:1 ratio.

It is unlikely that the apparently grouped classes of genes are coordinately regulated by a common regulatory mechanism analogous to a locus control region. This is proposed for two reasons. First, the three classes of genes reported in this work show extensively overlapping expression patterns (Wang, 1996) making it difficult to understand how members of different classes, under the control of a different regulatory program would display a similar expression pattern. Also arguing against a locus control region are the observations made by Soares et al. (1987) regarding the expression of an alpha-2u-globulin transgene. An alpha-2u-globulin transgene with 2.7 kb of upstream sequence demonstrated the expected expression patterns in transgenic mice, indicating that the sequences required for the correct expression of the transgene are in close proximity to the gene and not in a more distant locus control region. Additionally, even within a class of gene it is apparent that there is different developmental expression patterns between the various tissues expressing that type (Wang, 1996). It is unlikely therefore that the clustering of gene types defined by RFLPs is reflective of a similar tissue specific, or developmental expression pattern for the group of genes.

The observation of small groups of 3-5 alpha-2u-globulin genes in close proximity following genomic DNA Fiber FISH is also suggestive of gene clustering within the alpha-2u-globulin locus. The relationship if any of these groups of genes to the three classes identified by RFLPs, and to any common expression pattern is not known. It is possible, however, that these smaller groups of alpha-2u-globulin may result from the breakage of DNA at one of the stages of the Fiber FISH protocol, thus giving the appearance of smaller groups of genes within the alpha-2u-globulin locus. This possibility will be resolved as physical mapping in the region continues.

Regardless of whether there are groupings of genes within the locus or not, it does seem clear that the spacing between alpha-2u-globulin genes is somewhat variable (4.2.2). If smaller groups of genes within the cluster do exist, it will be of interest to determine if there is any functional significance to the groupings, for example- do grouped genes share a similar tissue-specific expression pattern?

5.4. ORIGIN OF THE ALPHA-2U-GLOBULIN REPEAT

The extensive homologies between alpha-2u-globulin genes, extending perhaps more than 12 kb in both the 5' and 3' flanking regions suggests that the original duplication event, regardless of its mechanism, must have involved a large DNA segment.

Four types of events have been suggested as possible mechanisms for the initial gene duplication required for the generation of a multigene family (Maeda and Smithies, 1986). These mechanisms include: Nonhomologous chromosomal breakage and reunion, homologous recombination between repeated elements, RNA-mediated transposition, and gene amplification through over-replication. At present is not possible to definitively determine which of these possible mechanisms can account for the original duplication event leading to the formation of the alpha-2u-globulin gene family. However, given the nature of the characterized alpha-2u-globulin gene sequences, it seems unlikely that the locus arose as the result of an RNA-mediated DNA duplication. All characterized genomic clones contain introns, upstream regulatory sequences, and lack remnants of a poly(A) tract. Duplication via an RNA intermediate is also most commonly associated with insertion of the duplicated gene at a distant site (Cross and Renkawitz, 1990). Additionally, there is no evidence that gene amplification through over-replication has occurred in the germ line of higher eukaryotes (Maeda and Smithies, 1986), making this mechanism a less likely candidate to explain the origin of the alpha-2u-globulin locus. Nonhomologous chromosomal breakage and reunion, and homologous recombination between repeated elements remain possibile explanations which could account for the origin and evolution of the current alpha-2u-globulin gene cluster.

5.5. EVOLUTIONARY RELATIONSHIP BETWEEN RAT ALPHA-2U-GLOBULIN AND MOUSE MUP GENES

Since the homologs, alpha-2u-globulin and MUP, are each present in multiple copies in the rat and mouse genomes respectively, it would be simplest to assume that the duplication events responsible for the multiple copies of these genes in each species occurred prior to their divergence, rather than independently in the two species. If following speciation, the duplicate genes within each species evolved independently of each other, the intra- and interspecific differences between these urinary proteins would be expected to be more or less the same (Arnheim, 1983).

The intraspecific divergences for alpha-2u-globulin and MUP are approximately 2-10 times lower than the interspecific divergence (Clark et al., 1984a). The rat alpha-2u-globulin genes differ from each other by less than 10% in their nucleotide sequences, as do the MUPs, however differences between these homologous gene families is about 20%. This phenomenon whereby intraspecific variability among members of a multigene family is much less than the interspecific differentiation between members of the homologous gene family is referred to as concerted evolution.

A set of identical duplicate genes in the ancestor of the mouse and rat would be expected to be subject to random independent mutational events in the lineages leading to the two species. Concerted evolution, involving genetic interactions between family members, occurring at random intervals results in the elimination of most spontaneous mutations. Eventually, these genetic interactions result in the fixation of a particular mutation in all of the members of the family within a species. Sequence homogenization within a family occurs because single base mutations are eliminated or spread to other members by genetic interactions between different genes of the family. Divergence between species results because the mutation and fixation processes are independent in the two lineages, therefore each species would become homogenous for different mutations. Concerted evolution can therefore account for both the high degree of intraspecies homogeneity, and the interspecies divergence observed between the rat alpha-2u-globulin genes and the mouse MUP genes.

Two types of genetic interaction are commonly discussed which would facilitate concerted evolution; unequal crossing-over and gene conversion. Unequal crossing-over between misaligned members of a multigene family can result in the expansion or contraction in the size of the family thus altering the frequency of individual members in the original array. Repeated cycles of unequal crossing-over results in the production of homogenous tandem arrays where all family members are derived from a small number of original family members (Ohta, 1980). Recombination leading to gene conversion can decrease the differences between family members by transfering a significant number of nucleotides from a donor gene to a recipient gene in one step. The clustering of all the alpha-2u-globulin genes in a region of less than 900 kb would be expected to facilitate these mechanisms of information exchange consistent with concerted evolution.

A requirement for gene families to evolve in a concerted fashion is that the rate of mutational divergence between family members is less than the rate of fixation (Arnheim, 1983). This ensures homogeneity among family members. The frequency of both gene conversion and unequal-crossing over, which are proposed to be the primary mechanisms operational in maintaining sequence homogeneity within families of repeated genes, has

been investigated in the yeast *Saccharomyces* cerevisiae.

Klein and Petes (1981) demonstrated a conversion frequency of 4% at the *LEU2* locus in *Saccharomyces cerevisiae*. This frequency is clearly higher than the frequency of mutation. As a result, gene conversion, at least in the context of the yeast genome, should be sufficient to maintain homogeneity of certain classes of repeated genes. The frequency of gene conversion may depend on a number of factors including: specific nucleotide sequences, the size of the repeated gene, and the distance between repeated units (Klein and Petes, 1981). Consequently, gene conversion may not provide a sufficient homogenizing force to remove sequence diversity generated by mutation at all repeated loci. The frequency of unequal crossing-over at the rDNA locus in *Saccharomyces cerevisiae* has been investigated (Petes, 1980). Petes (1980) demonstrated unequal crossing-over in approximately 10% of tetrads. Due to the detection limits of the assay, this should be considered a conservative estimate, with the true frequency of unequal crossing-over likely being much greater than 10%. It appears then, that in many cases, the combined frequency of gene conversion and unequal crossing-over should be sufficient to counter-balance the effects of mutational divergence between individual members of repetitive families.

The 18-20 copies of the alpha-2u-globulin gene estimated to be present in the rat genome (Kurtz, 1981; this work) is approximately equivalent to the number of MUP genes in each of the two groups (Clark et al., 1984b). The apparent difference in the size of the alpha-2u-globulin family compared to the MUP gene family may reflect the independent amplification of these homologous genes in the two species, but more likely represents the result of multiple expansions and contractions brought about by unequal crossing over.

The 30 late chorion protein genes in *Bombyx mori* have a genomic organization resembling that of the MUP genes. The 30 genes are organized as 15 tandem pairs, with the genes in each pair divergently oriented. The genes in each pair have a sequence identity of about 91% (Xiong et al., 1988). It has been suggested that this rather complex repeating unit is more likely to have arisen out of a pre-existing tandem array than from a series of fifteen isolated duplications (Graham, 1995). Given the structure of the MUP locus, a similar argument can be envoked to explain the origin of its present organization.

The organization of the rat alpha-2u-globulin genes, predominantly as head to tail tandem repeats, may more closely resemble the organization of the genes in the common ancestor of mice and rats. An inversion event in the ancestral array of genes in the lineage leading to the mouse followed by repeated cycles of unequal crossing-over could result in the replacement of

any pre-existing tandem array of genes with an array comprised of divergently oriented gene pairs characteristic of the MUP locus.

However, the occurrence of several genes in an inverted orientation throughout the rat alpha-2u-globulin locus suggests that one of these regions in the present day rat genome may have served as the genetic substrate in the events leading to the organization of the MUP locus. Alternatively the events which have produced the localized inverted regions in the rat alpha-2u-globulin locus may have post-dated the divergence of the mouse and rat. What is evident is that within the present day rat alpha-2u-globulin locus, gene arrangements exist that could be duplicated and expanded to generate a gene array resembling that of the present day MUP gene cluster.

The presence of oppositely oriented short repeats could be responsible for the localized gene inversions that disrupt the otherwise tandem organization of alpha-2u-globulin genes as presented in Figure 4.7 (Jeffreys and Harris, 1982). Similarily, these types of sequence elements could have been used in the replacement of a tandem array by the inverted gene pair in the mouse lineage. The mouse repetitive elements R and B1 have been demonstrated to be part of the 45 kb pallindrome which forms the predominant organizational unit at the MUP locus (Bishop et al., 1985). The potential involement of these elements in the evolution of the MUP locus has not been investigated.

In the rat a repetitive element has also been mapped to the upstream region of the alpha-2u-globulin gene represented by clones pL 2, 8, 9, 10, however, no evidence for a repetitive element at the corresponding position of clone pL1. Additionally, Yamamoto et al. (1989) demonstrated the presence of a LINE 1 element in the sixth intron of an isolated alpha-2u-globulin genomic clone. The distribution of repetitive elements within the alpha-2u-globulin locus warrants a more in depth analysis so that the significance, if any, of their role in the evolution of the alpha-2u-globulin locus can be determined.

5.6. IMPLICATIONS OF ALPHA-2U-GLOBULIN GENE EVOLUTION ON REPRODUCTIVE ISOLATION

Speciation depends on the evolution of effective prezygotic and/or postzygotic isolation mechanisms. Prezygotic reproductive isolation, especially mating discrimination, may be the primary cause of speciation in many animal taxa (Campbell, 1993). Among prezygotic isolation mechanisms, olfactory communication (i.e. through pheromones) has been demonstrated to play a major role in many aspects of mammalian adaption and speciation including involvement in behaviors such as intra- and

interspecies identification, social structure and organization, territory marking and alarm signalling (Menzies et al., 1992).

Assortative mating has evolutionary consequences on the genetic structure of populations and the processes of speciation particularily as a prezygotic reproductive isolating mechanism (Nevo et al., 1976). Examples of both positive and negative assortive mating have been observed. For example, in the mouse, positive assortative mating has been suggested to reinforce sexual isolation between subspecies, whereas negative assortive mating has been found in different strains within subspecies and is believed to reinforce outbreeding and increase heterozygosity (Yanai and McCleam, 1973).

The positive assortive mating behavior of the parapatric chromosomal forms of the mole rat, *Spalax ehrenbergi*, has been investigated in some detail (Nevo et al., 1976; Nevo and Heth, 1976; Menzies et al., 1992). The chromosomal forms are apparently in the final stages of speciation, as evidenced by the narrowing of hybrid zones separating them. Since speciation of the different chromosomal forms seems incomplete, prezygotic reproductive isolation mechanisms would complement the cytological barriers inherent in matings between individuals of the different chromosomal forms.

Estrus females from the various parapatric chromosomal forms have been shown to demonstrate a significant mating preference for males of their own chromosomal form (Nevo et al., 1976). It was further demonstrated that components of male mole rat urine, possibly pheromones, were responsible for the ability of estrus females to discriminate conspecific, or homochromosomal males (Nevo and Heth, 1976). This olfactory discrimination was only observed with estrus females, and was therefore speculated to play a role in the speciation of the chromosomal forms as a prezygotic reproductive isolation mechanism, complementing the chromosomal incompatibilities and contributing to the final stages of speciation.

Similar examples of olfactory discrimination have been noted in a number of rodent species (Doty, 1986) and insects (Coyne et al., 1994; Tumlinson et al., 1974) suggesting that pheromonal cues could be instrumental in allowing individuals of closely related species to distinguish each other and may serve as a widely used component of reproductive isolation mechanisms during speciation.

The demonstrated involvement of olfactory cues in a significant number of behaviors suggests individuals within a population, or species may synthesize a number of these chemical cues. It has been speculated that a number of these pheromones are bound/transported by accessory proteins

and that the protein-pheromone complex is taken up into the vomeronasal organ where an effect is initiated (Bacchini et al., 1992; Menzies et al., 1992). The mouse MUPs and rat alpha-2u-globulin are proposed to function in such a fashion.

The effectiveness of a chemical cue would then depend in part on the association with its' carrier protein. Speciation may be facilitated then by the rapid evolution of proteins that carry chemical cues playing a role in the reproductive isolation of the organisms. This hypothesis can be used to help explain the apparently rapid evolution of the mouse MUP and rat alpha-2u-globulin genes.

If we take the time since the rat-mouse divergence to be 30 million years, Clark et al. (1984a) have determined the unit evolutionary periods (the time required to produce a 1% sequence divergence) for the MUP/alpha-2u-globulin genes to be 0.33 million years and 1.31 million years for silent sites and replacement sites respectively. These values indicate that these gene families are evolving at a much more rapid rate than many other multigene families. For example, the human β-globin gene family has an estimated unit evolutionary period of about 10 million years (Efstratiadis et al., 1980).

It is possible that the rapid evolution of the MUP and alpha-2u-globulin genes is reflective of a role, through their pheromone transport capabilities, in mice and rats respectively, in establishing and maintaining intra- and interspecies identification codes and social structure and organization required for reproductive isolation.

If rat alpha-2u-globulin and the mouse MUPs do play such a role, it would be expected that there would be differences in these genes and/or their protein isoforms between inbred strains, or species. This appears to be the case, at least with the mouse MUP genes, where it has been investigated. Clissold and Bishop (1982) demonstrated inter-line variation in the isoelectric focussing profiles of both the urinary MUPs, and the unprocessed in vitro translation products obtained from hepatic mRNA. Additionally, minor inter-line differences were detected in mouse genomic Southern blots hybridized with a MUP cDNA probe. Most of the differences, in both the isoelectric focussing profiles and the Southern hybridizations seem to be quantitative rather than qualitative. That is, changes in the number of existing genes within the cluster rather than the appearance of novel gene types. This may indicate that different mouse lines may use a different subset of MUP proteins to effect their function and that changes in gene numbers and/or types are possible over relatively short evolutionary periods.

The observations that sexual isolation between species may sometimes have a simple genetic basis (Coyne et al., 1994; Thumlinson et al., 1974)

coupled with the proposed function and observed rapid evolution of the MUP and alpha-2u-globulin proteins leaves open the possibility that these proteins may contribute to prezygotic reproductive isolation mechanisms.

5.7. FUTURE STUDIES

The alpha-2u-globulin gene family provides an excellent system for the study of tissue, hormonal, and developmental control of gene expression. However the large size of the gene family, and the significant nucleotide homology between family members have complicated establishing the regulatory mechanisms operating on this family. A complete understanding of the organization of the locus is viewed as a prerequisite to comprehending the inherent complexities of the regulation of this gene family. The work reported in this thesis provides the first in depth analysis of the organization of the rat alpha-2u-globulin locus.

The cloning of the alpha-2u-globulin locus in large fragments containing multiple genes, as has been initiated in the present study, should hasten the identification of regulatory sequences responsible for the characteristic expression patterns family members are subjected to. Since several lines of evidence suggest that locus control regions are unlikely to be operating within the alpha-2u-globulin locus, efforts should be focussed on individual genes and their flanking sequences to determine the mechanisms responsible for the diverse regulation of family members. In this regard, efforts should be directed toward the continued construction of a contig covering the locus using BAC clones and PAC/P1 clones, if and when they become available.

The availability of a contig covering the alpha-2u-globulin locus would also allow for a more reliable analysis of the the complete organization of the locus using Fiber FISH. The use of BACs/PACs as targets in Fiber FISH studies, as oppossed to total genomic DNA, would permit physical measurements similar to those obtained from P1 Fiber FISH images. This would alleviate any uncertainties associated with piecing together information from profiles obtained from total genomic DNA, since DNA counter-stainable fibers of known lengths would be used as hybridization targets.

The construction of a contig, and the availability of alpha-2u-globulin genes known to originate from adjacent regions of the genome that would result, will permit the systematic testing both *in vitro* and *in vivo* of the possibility that genes from the same region of the locus are subject to similar tissue, developmental, and hormonal regulatory programs. This type of analysis could be achieved by introducing the gene(s) to be evaluated into a tissue culture cell line, or into mice, to examine the expression properties. Alternatively, with access to the nucleotide sequences of the entire

complement of alpha-2u-globulin genes, it may be possible to generate gene specific oligonucleotides from regions displaying divergence. This would permit the discrimination between members of the gene family, and thus the accumulation of data regarding the expression patterns of specific genes within the cluster.

An analysis for repeated sequences, or elements that may have played a role in the original amplification of the alpha-2u-globulin genes, or that could account for the inverted genes within the array should provide additional information which would aid in the understanding of the organizatioal evolution of the alpha-2u-globulin locus. The relationship between the C1 and C2 elements of the MUP pallindrome and the regions of inverted symmetry that flank the alpha-2u-globulin gene should be probed in an effort to establish the events that have lead to the contrasting organization of these homologous gene families in the mouse and rat.

CHAPTER 6.
REFERENCES

Abidi, F.E., Wada, M., Little, R.D., and Schlessinger, D. (1990). Yeast artificial chromosomes containing human Xq24-Xq28 DNA: Library construction and representation of probe sequences. *Genomics* **7:** 363-376.

Adams, M.D., Soares, M.B., Kerlavage, A.R., Fields, C., and Venter, J.C. (1993). Rapid cDNA sequencing (expressed sequence tags) from a directionally cloned human infant brain cDNA library. *Nature Genet.* **4:** 373-380.

Addison, W.R., and Kurtz, D.T. (1986). Nucleotide sequences required for the regulation of a rat α_{2u}-globulin gene by glucocorticoids. *Mol. Cell. Biol.* **6:** 2334-2346.

Addison, W.R., and Kurtz, D.T. (1989). Identification of nuclear proteins that bind to the glucocorticoid regulatory region of a rat alpha-2u-globulin gene. *J. Biol. Chem.* **264:** 21891-21895.

Alden, C.L., Kanerva, R.L., Ridder, G., and Stone, L.C. (1984). The pathogenesis of the nephrotoxicity of volatile hydrocarbons in the male rat. *Adv. Mod. Environ. Toxicol.* **7:** 107-120.

Anand, R., Riley, J.H., Butler, R., Smith, J.C., and Markham, A.F. (1990). A 3.5 genome equivalent multi access YAC library: Construction, characterization, screening and storage. *Nucl. Acids Res.* **18:** 1951-1956.

Arnheim, N. (1983). Concerted evolution of multigene families. *In* "Evolution of Genes and Proteins". pp. 38-61, Sinauer Associates, Sunderland.

Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A., and Struhl, K. (1995). *In* "Current Protocols in Molecular Biology". John Wiley & Sons Inc., New York.

Bacchini, A., Gaetani, E., and Cavaggioni, A. (1992). Pheromone binding proteins of the mouse, *Mus musculus*. *Experientia* **48:** 419-421.

Baltimore, D. (1981). Gene conversion: Some implications for immunoglobulin genes. *Cell* **24:** 592-594.

Barletta, C., Batticane, N., Ragusa, R.M., Leube, R., Peschle, C., and Romano, V. (1990). Subchromosomal localization of two human cytokeratin genes (KRT4 and KRT15) by in situ hybridization. *Cytogenet. Cell Genet.* **54:** 148-150.

Baumann, H., and Held, W.A. (1981). Biosynthesis and hormone-regulated expression of secretory glycoproteins in rat liver and hepatoma cells. Effect of glucocorticoids and ininflammation. *J. Biol. Chem.* **256:** 10145-10155.

Baumann, H., Firestone, G.L., Burgess, T.L., Gross, K.W., Yamamoto, K.R., and Held, W.A. (1983). Dexamethasone regulation of alpha-1-acid glycoprotein and other acute phase reactants in rat liver and hepatoma cells. *J. Biol. Chem.* **258:** 563-570.

Baxendale, S., MacDonald, M.E., Mott, R., Francis, F., Lin, C., Kirby, S.F., James, M., Zehetner, G., Hummerich, H., Valdes, J., Collins, F., Deaven, L., Gusella, J., Lehrach, H., and Bates, G. (1993). *Nature Genet.* **4:** 181-186.

Beckmann, J.S., and Weber, J.L. (1992). Survey of human and rat microsatellites. *Genomics* **12:** 627-631.

Bellanne-Chantelot, C., Lacroix, B., Ougen, P., Billault, A., Beaufils, S., Bertrand, S., Georges, I., Gilbert, F., Gros, I., Lucotte, G. (1992). Mapping the whole human genome by fingerprinting yeast artificial chromosomes. *Cell* **70:** 1059-1068.

Bengtsson, U., Altherr, M.R., Wasmuth, J.J., and Winokur, S.T. (1994). High resolution fluorescence in situ hybridization to linearly extended DNA visually maps a tandem repeat associated with facioscapulohumeral muscular dystrophy immediately adjacent to the telomere of 4q. *Hum. Mol. Genet.* **3:** 1801-1805.

Berger, S., Zellmer, D.M., Kloppstech, K., Richter, G., Dillard, W.L., and Schweiger, H.G. (1978). Alternating polarity in rDNA genes. *Cell Biol. Int. Reports* **2:** 41-50.

Bernstein, L.B., Manser, T., and Weiner, A.M. (1985). Human U1 small nuclear RNA genes: Extensive conservation of flanking sequences suggests cycles of gene amplification and transposition. *Mol. Cell. Biol.* **5:** 2159-2171.

Bickmore, W.A., and Sumner, A.T. (1989). Mammalian chromosome banding- an expression of genome organization. *TIGS.* **5:** 144-148.

Birnboim, H.C., and Doly, J. (1979). A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nuc. Acid Res.* **7:** 1513-1521.

Birren, B., and Lai, E. (1993). Pulsed Field Gel Electrophoresis: A Practical Guide. Academic Press, Inc., New York.

Bishop, J.O., Clark, A.J., Clissold, P.M., Hainey, S., and Francke, U. (1982). Two main groups of mouse major urinary protein genes, both largely located on chromosome 4. *EMBO J.* **1:** 615-620.

Bishop, J.O., Selman, G.G., Hickman, J., Black, L., saunders, R.D.P., and Clark, A.J. (1985). The 45 kb unit of major urinary protein gene organization is a gigantic imperfect palindrome. *Mol. Cell. Biol.* **5:** 1591-1600.

Blackburn, E.H. (1984). Telomeres: do the ends justify the means? *Cell* **37:** 7-8.

Bloomhoff, R., Gren, M.H., Berg, T., and Norum, K.R. (1990). Transport and storage of vitamin A. *Science* **250:** 399-404.

Bocskei, Z., Groom, C.R., Flower, D.R., Wright, C.E., Phillips, S.E.V., Cavaggioni, A., Findlay, J.B.C., and North, A.C.T. (1992). Pheromone binding to two rodent urinary proteins revealed by X-ray crystallography. *Nature* **360:** 186-188.

Booth, W.D. (1984). Sexual dimorphism involving steroidal pheromones and their binding proteins in the submaxillary salivary gland of the Gottingen miniature pig. *J. Endocrin.* **100:** 195-202.

Booth, W.D., and White, C.A. (1988). The isolation, purification and some properties of pheromaxein, the pheromonal steroid-binding protein in porcine submaxillary glands and saliva. *J. Endocrin.* **118:** 47-57.

Borghoff, S.J., Short, B.G., and Swenberg, J.A. (1990). Biochemical mechanisms and pathobiology of α_{2u}-globulin nephropathy. *Ann. Rev. Pharmacol. Toxicol.* **30:** 349-367.

Botstein, D., White, R.L., Skolnick, M., and Davis, R.W. (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphism. *Am. J. Hum. Genet.* **32:** 314-331.

Britten, R.J. and Kohne, D.E. (1969). Repetition of nucleotide sequences in chromosomal DNA. *In* "Handbook of Molecular Cytology". pp. 21-36, North-Holland Publishing Company, Amsterdam.

Britten, R.Y., Baron, W.F., Stout, D.B., and Davidson, E.H. (1988). Sources and evolution of human *Alu* repeated sequences. *Proc. Natl. Acad. Sci. USA* **85:** 4770-4774.

Brook, J.D., McCurrach, M., Harley, H.G., Buckler, A.J., Church, D., Aburatani, H., Hunter, K., Stanton, V.P., Thirion, J.P., Hudson, T. (1992). Molecular basis of myotonic dystrophy: Expansion of a trinucleotide (CTG) repeat at the 3' end of a transcript encoding a protein kinase family member. *Cell* **68:** 799-808.

Brooks, D.E., Means, A.R., Wright, E.J., Singh, S.P., and Tiver, K.K. (1986). Molecular cloning of the cDNA for two majorandrogen-dependent secretory proteins of 18.5 kilodaltons synthesized by the rat epididymis. *J. Biol. Chem.* **261:** 4956-4961.

Brownstein, B.H., Silverman, G.A., Little, R.D., Burke, D.T., Korsmeyer, S.J., Schlessinger, D., and Olson, M.V. (1989). Isolation of single-copy human genes from a library of yeast artificial chromosome clones. *Science* **244:** 1348-1351.

Buckle, V.J., and Kearney, L. (1994). New methods in cytogenetics. *Curr. Opin. Genet. Dev.* **4:** 374-382.

Bundgaard, J.R., Sengelov, H., Borregaard, N., and Kjeldsen, L. (1994). Molecular cloning and expression of a cDNA encoding NGAL: A lipocalin expressed in human neutrophils. *Biochem. Biophys. Res. Comm.* **202:** 1468-1475.

Burke, D.T., Carle, G.F., and Olson, M.V. (1987). Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors. *Science* **236:** 806-812.

Burton, F.H., Loeb, D.D., Voliva, C.F., Martin, S.L., Edgell, M.H., and Hutchinson, C.A. (1986). Conservation throughout mammalia and extensive protein-coding capacity of the highly repeated DNA long interspersed sequence one. *J. Mol. Biol.* **187:** 291-304.

Callan, H.G. and Lloyd, L. (1960). Lampbrush chromosomes of crested newts *Triturus cristatus* (Laurenti). *Phil Trans. R. Soc. B* **243:** 135-219.

Campbell, N.A. (1993). Biology. Benjamin/Cummings Publishing Company, Inc.

Card, C.O., Morris, G.F., Brown, D.T., and Marzluff, W.F. (1982). Sea urchin small nuclear RNA genes are organized in distinct tandemly repeating units. *Nucl. Acids Res.* **10:** 7677-7688.

Cavaggioni, A., Findlay, J.B.C., and Tirindelli, R. (1990). Ligand binding characteristics of homologous rat and mouse urinary proteins and pyrazine binding protein of the calf. *Comp. Biochem. Physiol.* **96B:** 513-520.

Chainani, M., Sampsell, B., and Elliott, R.W. (1991). Localization of the gene for plasma retinol binding protein to the distal half of mouse chromosome 19. *Genomics* **9:** 376-379.

Chan, G.C-K., Hess, P., Meenakshi, T., Carlstedt-Duke, J., Gustafsson, J-A., and Payvar, F. (1991). Delayed secondary glucocorticoid response elements. *J. Biol. Chem.* **266:** 22634-22644.

Chan, G.C., Hess, P., Meenakashi, T., Carstedt-Duke, J., Gustafsson, J.A., and Payvar, F. (1991). Delayed secondary glucocorticoid response elements. Unusual nucleotide motifs specify glucocorticoid receptor binding to transcribed regions of alpha-2u-globulin DNA. *J. Biol. Chem.* **266:** 22634-22644.

Chan, P., Simon-Chazottes, D., Mattei, M.G., Guenet, J.L., and Salier, J.P. (1994). Comparative mapping of lipocalin genes in human and mouse: The four genes for complement C8γ chain, prostaglandin-D-synthetase, oncogene-24P3, and progestagen-associated endometrial protein map to HSA9 and MMU2. *Genomics* **23:** 145-150.

Charbonneau, M., Strasser, J., Borghoff, S.J., and Swenberg, J.A. (1988). In vitro hydrolysis of [14C]-α_{2u}-globulin (α 2u) isolated from male rat kidney. *Toxicologist* **8:** 135.

Charbonneau, M., Strasser, J., Lock, E.A., Turner, M.J., and Swenberg, J.A. (1989). Involvement of reversible binding to α_{2u}-globulin in 1,4-dichlorobenzene-induced nephrotoxicity. *Toxicol. Appl. Pharmacol.* **99:** 122-132.

Chatterjee, B., Demyan, W.F., and Roy, A.K. (1983). Interesting role of thyroxine and growth hormone in the hepatic synthesis of α 2u globulin and its messenger RNA. *J. Biol. Chem.* **258:** 688-692.

Chen, C-L.C., and Feigelson, P. (1978a). Hormonal induction of α_{2u} globulin synthesis in isolated rat hepatocytes. *Biochemistry* **17:** 5308-5313.

Chen, C-L.C., and Feigelson, P. (1978b). Glucocorticoid induction of α_{2u} globulin protein synthesis and its mRNA in rat hepatocytes *in vitro*. *J. Biol. Chem.* **253:** 7880-7885.

Chen, C-L.C., and Feigelson, P. (1979). Cyclohexamide inhibition of hormonal induction of α_{2u} globulin mRNA. *Proc. Natl. Acad. Sci. USA* **76:** 2669-2673.

Childs, G., Maxson, R., Cohn, R.H., and Kedes, L. (1981). Orphons: Dispersed genetic elements derived from tandem repetitive genes of eukaryotes. *Cell* **23:** 651-663.

Chipman, R.K., and Albrecht, E.D. (1974). The relationship of the male preputial gland to the acceleration of oestrus in the laboratory mouse. *J. Reprod. Fert.* **38:** 91-96.

Choy, H.A., McLaughlin, M., and Feigelson, P. (1989). Nuclear factors from expressing tissues interact *in vitro* with a rat alpha-2u-globulin gene intron. *J. Biol. Chem.* **264:** 1754-1759.

Clark, A.J., Chave-Cox, A., Ma, X., and Bishop, J.O. (1985b). Analysis of mouse major urinary protein genes: Variation between the exonic sequences of Group 1 genes and a comparison with an active gene outwith Group 1 both suggest that gene conversion has occurred between MUP genes. *EMBO J.* **4:** 3167-3171.

Clark, A.J., Clissold, P.M., and Bishop, J.O. (1982). Variation between mouse major urinary protein genes isolated from a single inbred line. *Gene* **18:** 221-230.

Clark, J.A., Clissold, P.M., Shawi, R.A., Beattie, P., and Bishop, J. (1984a). Structure of the mouse major urinary protein genes: Different splicing configurations in the 3'-non-coding region. *EMBO J.* **3:** 1045-1052.

Clark, J.A., Ghazal, P., Bingham, R.W., Barrett, D., and Bishop, J.O. (1985a). Sequence structures of a mouse major urinary protein gene and pseudogene compared. *EMBO J.* **4:** 3159-3165.

Clark, J.A., Hickman, J., and Bishop, J. (1984b). A 45-kb DNA domain with two divergently oriented genes is the unit of organisation of the murine major urinary protein genes. *EMBO J.* **3:** 2055-2064.

Clark-Adams, C.D., Norris, D., Osley, M.A., Fassler, J.S., and Winston, F. (1988). Changes in histone gene dosage alter transcription in yeast. *Genes Dev.* **2:** 150-159.

Claude, A. and Potter, T.S. (1943). Isolation of chromatin threads from the resting nucleus of leukemic cells. *J. Exp. Med.* **77:** 345-354.

Clissold, P.M., and Bishop, J.O. (1982). Variation in the mouse urinary protein (MUP) genes and the MUP gene products within and between inbred lines. *Gene* **18:** 211-220.

Colby, C., and Williams, S.M. (1993). The distribution and spreading of rare variants in the histone multigene family of *Drosophila melanogaster*. *Genetics* **135:** 127-133.

Cooper, D.N. (1994). Mapping the human genome. *Ann. Genet.* **37:** 101-116.

Copeland, N.G., Jenkins, N.A., Gilbert, D.J., Eppig, J.T., Maltais, L.J., Miller, J.C., Dietrich, W.F., Weaver, A., Lincoln, S.E., Steen, R.G., Stein, L.D., Nadeau, J.H., and Lander, E.S. (1993). A genetic linkage map of the mouse: Current applications and future prospects. *Science* **262:** 57-66.

Coulson, A., Waterson, R., Kiff, J., Sulston, J., and Kohara, Y. (1988). Genome linking with yeast artificial chromosomes. *Nature* **335:** 184-186.

Cowley, J.J. (1978). Olfaction and the development of sexual behavior. *In* "Biological determinants of sexual behaviour". pp. 87-95, John Wiley & Sons, Chichester.

Coyne, J.A., Crittenden, A.P., and Mah, K. (1994). Genetics of a pheromonal difference contributing to reproductive isolation in Drosophila. *Science* **265:** 1461-1464.

Craig, I.W. (1994). Organization of the human genome. *J. Inher. Metab. Dis.* **17:** 391-402.

Cross, M., and Renkawitz, R. (1990). Repetitive sequence involvement in the duplication and divergence of mouse lysozyme genes. *EMBO J.* **9:** 1283-1288.

Cross, M., and Renkawitz, R. (1990). repetitive sequence involvement in the duplication and divergence of mouse lysozyme genes. *EMBO J.* **9:** 1283-1288.

Cutler, W.B., Preti, G., Krieger, A.M., Huggins, G.R., and Lawley, H.J. (1986). *Horm. Behav.* **20:** 463-473.

D'Andrea, R.J., Coles, L.S., Lesnikowski, C., Tabe, L., and Wells, J.R.E. (1985). Chromosomal organization of chicken histone genes: Preferred associations and inverted duplications. *Mol. Cell. Biol.* **5:** 3108-3115.

Dahlberg, J., and Lund, E. (1988). The genes and transcription of the major small nuclear RNAs. *In* "Structure and Function of Major and Minor Small Nuclear Ribonucleoprotein Particles". pp. 38-70, Springer-Verlag, Berlin.

Dauwerse, J.G., Wiegant, J., Raap, A.K., Breuining, M.H., and van Ommen, G.J.B. (1992). Multiple colors by fluorescence *in situ* hybridization using ratio-labeled DNA probes creates a molecular karyotype. *Hum. Mol. Genet.* **1:** 593-598.

De Angioletti, M., Lacerra, G., Castaldo, C., Cutolo, R., De Bonis, C., Buonanno, G, and Carestia, C. (1992). Alpha alpha alpha alpha (anti 3.7 type II): a new alpha-globin gene rearrangement suggesting that the alpha-globin gene duplication could be caused by intrachromosomal recombination. *Hum. Genet.* **89:** 37-41.

Deatrick, J., Daly, M., Randsholt, N.B., and Brock, H.W. (1991). The complex genetic locus *polyhomeotic* in *Drosophila melanogaster* potentially encodes two homologous zinc finger proteins. *Gene* **105:** 185-195.

den Dunnen, J.T., Moormann, R.J.M., Lubsen, N.H., and Schoenmakers, J.G.G. (1986). Concerted and divergent evolution within the rat γ-crystallin gene family. *J. Mol. Biol.* **189:** 37-46.

Dente, L., Pizza, M.G., Metspalu, A., and Cortese, R. (1987). Structure and expression of the genes coding for human α1-acid glycoprotein. *EMBO J.* **6:** 2289-2296.

Diaz, M.O., Pomykala, H.M., Bohlander, S.K., Maltepe, E., Malik, K., Brownstein, B., and Olopade, O., I. (1994). Structure of the human type-I interferon gene cluster determined from a YAC clone contig. *Genomics* **22:** 540-552.

Dietrich, D.R., and Swenberg, J.A. (1991). The presence of α_{2u}-globulin is necessary for d-limonene promotion of male rat kidney tumors. *Cancer Res.* **51:** 3512-3521.

Dietrich, W., Katz, H., Lincoln, S.E., Shin, H-S., Friedman, J., Dracopoli, N.C., and Lander, E.S. (1992). A genetic map of the mouse suitable for typing intraspecific crosses. *Genetics* **131:** 423-447.

DiNocera, P.P., and Sakaki, Y. (1990). LINEs: A superfamily of retrotransposable ubiquitous DNA elements. *TIGS.* **6:** 29-30.

Doglio, A., Dani, C., Grimaldi, P., and Ailhaud, G. (1989). Growth hormone stimulates c-fos gene expression by means of protein kinase C without increasing inositol lipid turnover. *Proc. Natl. Acad. Sci. USA* **86:** 1148-1152.

Dolan, K.P., Unterman, R., McLaughlin, M., Nakhasi, H.L., Lynch, K.R., and Feigelson, P. (1982). The structure and expression of very closely related members of the α 2u globulin gene family. *J. Biol. Chem.* **257:** 13527-13534.

Dolan, K.R., Reinke, R., Lynch, K.R., McLaughlin, M., Unterman, R., Nakhasi, H.K., Laperche, Y., and Feigelson, P. (1983). The application of recombinant techniques to the study of the control of alpha 2u globulin gene expression. *In* "Biochemical Actions of Hormones X". pp. 1-19, Academic Press Inc., New York.

Donis-Keller, H., Green, P., Helms, C., Cartinhour, S., Wiffenbach, B., Stephens, K., Keith, T.P., Bowden, D.W., Smith, D.R., Lander, E.S., Botstein, D., Akots, G., Rediker, K.S., Gravius, T., Brown, V.A., Rising, M.B., Parker, C., Powers, J.A., Watt, D.E., Kauffman, E.R., Bricker, A., Phipps, P., Muller-Kahle, H., Fulton, T.R., Ng, S., Schumm, J.W., Braman, J.C., Knowlton, R.G., Barker, D.F., Crooks, S.M., Lincoln, S.E., Daly, M.J., and Abrahamson, J. (1987). A genetic linkage map of the human genome. *Cell* **51:** 319-337.

Doty, R.L. (1986). Odor-guided behavior in mammals. *Experientia* **42:** 257-271.

Dover, G.A. (1982). Molecular drive: a cohesive mode of species evolution. *Nature* **299:** 111-117.

Dover, G., Brown, S., Coen, E., Dallas, J., Strachan, T., and Trick, M. (1982). The dynamics of genome evolution and species

differentiation. *In*"Genome Evolution". pp. 343-372, Academic Press Inc., New York.

Drayna, D.T., McLean, J.W., Wion, K.L., Trent, J.M., Drabkin, H.A., and Lawn, R.M. (1987). Human apolipoprotein D gene: Gene sequence, chromosome localization, and homology to the alpha-2u-globulin superfamily. *DNA* **6:** 199-204.

Drickamer, K., Kwoh, T.J., and Kurtz, D.T. (1981). Amino acid sequence of the precursor of rat liver α_{2u}-globulin. *J. Biol. Chem.* **256:** 3634-3636.

Drouin, G., Sevigny, J-M, McLaren, I.A., Hofman, J.D., and Doolittle, W.F. (1992). Variable arrangement of 5 S ribosomal genes within the ribosomal DNA repeats of arthropods. *Mol. Biol. Evol.* **9:** 826-835.

Eden, S. (1979). Age- and sex-related differences in episodic growth hormone secretion in the rat. *Endocrinology* **105:** 555-560.

Efstratiadis, A., Posakony, J.W., Maniatis, T., Lawn, R., O'Connell, C., Spritz, R.A., DeRiel, J.K., Forget, B.G., Weissman, S.M., Slightom, J.L., Blechl, A.E., Smithies, O. (1980). The structure and evolution of the human β-globin gene family. *Cell* **21:** 653-668.

Enberg, B., Hulthen, A., Moller, C., Norstedt, G, and Francis, S.M. (1994). Growth hormone (GH) regulation of a rat serine protease inhibitor fusion gene in cells transfected with GH receptor cDNA. *J. Mol. Endocrin.* **12:** 39-46.

Evans, B.A., Drinkwater, C.C., and Richards, R.I. (1987). Mouse glandular kallikrein genes. structure and partial sequence analysis of the kallikrein gene locus. *J. Biol. Chem.* **262:** 8027-8034.

Evans, B.A., Yun, Z.X., Close, J.A., Tregear, G.W., Kitamura, N., Nakanishi, S., Callen, D.F., Baker, E., Hyland, V.J., Sutherland, G.R., and Richards, R.I. (1988). Structure and chromosomal localization of the human renal kallikren gene. *Biochemistry* **27:** 1324-3129.

Evans, T., Felsenfeld, G., and Reitman, M. (1990). Control of globin gene transcription. *Ann. Rev. Cell Biol.* **6:** 95-124.

Feigelson, P., and Kurtz, D.T. (1978). Hormonal modulation of specific messenger RNA species in normal and neoplastic rat liver. *Adv. Enzym.* **47:** 275-312.

Feinberg, A.P., and Vogelstein, B. (1983). A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Anal. Biochem.* **132:** 6-13.

Felicioli, A., Ganni, M., Garibotti, M., and Pelosi, P. (1993). Multiple types and forms of odorant-binding proteins in the old-world porcupine *Hystrix cristata*. *Comp. Biochem. Physiol.* **105B:** 755-784.

Feulgen, R. and Rossenbeck, H. (1924). Mickroskopisch-chemischer Nachweis einer Nucleinsaure vom Typus der Thymonucleinsaure und die daruf beruhende elektive Farbung von Zellkernen in mikroskopischen Proparaten. *Z. Physiol. Chem.* **135:** 203-247.

Fidlerova, H., Senger, G., Kost, M., Sanseau, P., and Sheer, D. (1994). Two simple procedures for releasing chromatin from routinely fixed cells for fluorescence *in situ* hybridization.

Filion, M., Sarafian, V., Lussier, M., Belanger, C., Lapointe, L., and Royal, A. (1994). Arrangement of a cluster of three mouse type I keratin genes expressed sequentially during esophageal-type epithelial cell differentiation. *Genomics* **24:** 303-310.

Finlayson, J.S., Asofsky, R., Potter, M., and Runner, C.C. (1965). Major urinary protein complex of normal mice: Origin. *Science* **149:** 981-982.

Flamm, W.G., Bond, H.E., Burr, H.E., and Bond, S.B. (1966). Satellite DNA isolated from mouse liver; some physical and metabolic properties. *Biochim. Biophys. Acta* **123:** 652-654.

Flower, D.R. (1994). The lipocalin protein family: a role in cell regulation. *FEBS Letts.* **354:** 7-11.

Flower, D.R. (1995). Multiple molecular recognition properties of the lipocalin protein family. *J. Mol. Recog.* **8:** 185-195.

Flower, D.R., North, A.C.T., and Attwood, T.K. (1993). Structural and sequence relationships in the lipocalins and related proteins. *Prot. Sci.* **2:** 753-761.

Flower, D.R., Sansom, C.E., Beck, M.E., and Attwood, T.K. (1995). The first prokaryotic lipocalins. *TIBS.* **20:** 498-499.

Fu, Y-H., Kuhl, D.P.A., Pizzutii, M., Pieretti, M, Sutcliffe, J.S., Richards, S., Verkerk, A., Holden, J., Fenwick, R., Warren, S.T. (1991). Variation

of the CGG repeat at the fragile X site results in genetic instability: Resolution of the Sherman Paradox. *Cell* **67:** 1047-1058.

Gall, J.G. and Callan, H.G. (1962). H^3 uridine incorporation in lampbrush chromosomes. *Proc. Natl. Acad. Sci. USA* **48:** 562-570.

Ganguly, M., Carnighan, R.H., and Westphal, U. (1967). Steroid-protein interactions. XIV. Interaction between human alpha-1-acid glycoprotein and progesterone. *Biochemistry* **6:** 2803-2814.

Gao, F., Endo, H., and Yamamoto, M. (1989). Length heterogeneity in rat salivary gland α2u globulin mRNAs: Multiple splice-acceptors and polyadenylation sites. *Nucl. Acids Res.* **17:** 4629-4636.

Garner, K.J., and Lingrel, J.B. (1988). Structural organization of the beta-globin locus of B-haplotype sheep. *Mol. Biol. Evol.* **5:** 134-140.

Garza, D., Ajioka, J.W., Burke, D.T., and Hartl, D.L. (1989). Mapping the Drosophila genome with yeast artificial chromosomes. *Science* **246:** 641-646.

Gasser, D.L., Sternberg, N.L., Pierce, J.C., Goldnersauve, A., Feng, H., Haq, A.K., Spies, T., Hunt, C., Buetow, K.H., and Chaplin, D.D. (1994). P1 and cosmid clones define the organization of 280 kb of the mouse H-2-complex containing the Cps-1 and Hsp70 loci. *Immunogenetics* **39:** 48-55.

Ghazal, P., Clark, A.J., and Bishop, J.O. (1985). Evolutionary amplification of a pseudogene. *Proc. Natl. Acad. Sci. USA* **82:** 4182-4185.

Gilbert, W. (1978). Why genes in pieces? *Nature* **271:** 501.

Glasgow, B.J., Abduragimov, A.R., Farahbakhsh, Z.T., Faull, K.F., and Hubbell, W.L. (1995). Tear lipocalins bind a broad array of lipid ligands. *Curr. Eye Res.* **14:** 363-372.

Godovac-Zimmermann, J. (1988). The structural motif of β-lactoglobulin and retinol-binding protein: A basic framework for binding and transport of small hydrophobic molecules? *TIBS.* **13:** 64-66.

Gorman, S.W., Roberts-Oehlschlager, S.L., Cullis, C.A., and Teasdale, R.D. (1992). *BioTechniques* **12:** 722-726.

Graham, G.J. (1995). Tandem genes and clustered genes. *J. Theor. Biol.* **175:** 71-87.

Granfornina, M.D., Sanchez, D., and Bastiani, M.J. (1995). Lazarillo, a new GPI-linked surface lipocalin is restricted to a subset of neurons in the grasshopper embryo. *Development* **121:** 123-134.

Grant, S., Jessee, J., Bloom, F., and Hanahan, D. (1990). Differential plasmid rescue from transgenic mouse DNAs into *Escherichia coli* methylation-restriction mutants. *Proc. Natl. Acad. Sci. USA* **87:** 4645-4649.

Grill, E., and Somerville, C. (1991). Construction and characterization of yeast artificial chromosome library of *Aradidopsis* which is suitable for chromosome walking. *Mol. Gen. Genet.* **226:** 484-490.

Guay, J-M., Huot, A., Gagnon, S., Tremblay, A., and Levesque, R.C. (1992). Physical and genetic mapping of cloned ribosomal DNA from *Toxoplasma gondii*: primary and secondary structure of the 5 S gene. *Gene* **114:** 165-171.

Gubits, R.M., Lynch, K.R., Kulkarni, A.B., Dolan, K.P., Gresik, E.W., Hollander, P., and Feigelson, P. (1984). Differential regulation of α 2u globulin gene expression in liver, lachrymal gland, and salivary gland. *J. Biol. Chem.* **259:** 12803-12809.

Gudas, L.J. (1994). Retinoids and vertebrate development. *J. Biol. Chem.* **269:** 15399-15402.

Gurland, G., Ashcom, G., Cochran, B.H., and Schwartz, J. (1990). Rapid events in growth hormone action. Induction of c-fos and c-jun transcription in 3T3-F442A preadipocytes. *Endocrinology* **127:** 3187-3195.

Guzman, P, and Ecker, J.R. (1988). Development of large DNA methods for plants: Molecular cloning of large segments of *Arabidopsis* and carrot DNA into yeast. *Nucl. Acids Res.* **16:** 11091-11105.

Haas, I.G., Simon-Chazottes, D., and Guenet, J.L. (1992). The gene coding for the immunoglobulin heavy chain binding protein BiP (*Hsce-70*) maps to mouse chromosome 2. *Mamm. Genome* **3:** 659-660.

Hamada, H. and Kakunaga, T. (1982). Potential Z-DNA forming sequences are highly dispersed in the human genome. *Nature* **298:** 396-398.

Hankeln, T., and Schmidt, E.R. (1991). The organization, localization and nucleotide sequence of the histone genes of the midge *Chironomus thummi*. *Chromosoma* **101:** 25-31.

Hansch, G.M. (1988). The homologous species restriction of the complement attack: Structure and function of the C8 binding protein. *Curr. Top. Miorobiol. Immunol.* **140:** 109-118.

Hard, G.C., Rodgers, I.S., Baetcke, K.P., Richards, W.L., McGaughy, R.E., and Valcovic, L.R. (1993). Hazard evaluation of the chemicals that cause accumulation of α_{2u}-globulin, hyaline droplet nephropathy, and tubule neoplasia in the kidneys of male rats. *Environ. Health Perspect.* **99:** 313-349.

Hastie, N., and Held, W. (1978). Analysis of mRNA populations by cDNA.mRNA hybrid-mediated inhibition of cell-free protein synthesis. *Proc. Natl. Acad. Sci. USA* **75:** 1217-1221.

Hastie, N.D., Held, W.A., and Toole, J.J. (1979). Multiple genes coding for the androgen-regulated major urinary proteins of the mouse. *Cell* **17:** 449-457.

Hattori, M., Kuhara, S., Takenaka, O., and Sakaki, Y. (1986). L1 family of repetitive DNA sequences in primates may be derived from a sequence encoding a reverse transcriptase-related protein. *Nature* **321:** 625-628.

Heiskanen, M., Hellsten, E., Kallioniemi, O., Makela, T.P., Alitalo, K., Peltonen, L., and Palotie, A. (1995). Visual mapping by Fiber-FISH. *Genomics* **30:** 31-36.

Heiskanen, M., Kallioniemi, O., and Palotie, A. (1996). Fiber-FISH: Experiences and a refined protocol. *Genet. Anal.* **12:** 179-184.

Heiskanen, M., Karhu, R., Hellsten, E., Peltonen, L., Kallioniemi, O.P., and Palotie, A. (1994). High resolution mapping using fluorescence *in situ* hybridization to extended DNA fibers prepared from agarose-embedded cells. *BioTechniques* **17:** 928-933.

Heitz, E. (1934). Die somatische heteropyknose bei *Drosophila melanogaster* und ihre genetische bedeutung. *Z. Zellforsch. Abt. Histochem.* **20:** 237-287.

Heng, H.H.Q., Squire, J., and Tsui, L.C. (1992). High resolution mapping of mammalian genes by *in situ* hybridization to free chromatin. *Proc. Natl. Acad. Sci. USA* **89:** 9509-9513.

Hentschel, C.C., and Birnsteil, M.L. (1981). The organization and expression of histone gene families. *Cell* **25:** 301-313.

Henzel, W.J., Rodriguez, H., Singer, A.G., Stults, J.T., Macrides, F., Agosta, W.C., and Niall, H. (1988). The primary structure of aphrodisin. *J. Biol. Chem.* **263:** 16682-16687.

Hess, P., Meenakashi, T., Chan, G.C., Carlstedt-Duke, J., Gustafsson, J.A., and Payvar, F. (1990). Purified glucocorticoid receptors bind selectively *in vitro* to a cloned DNA fragment that mediates a delayed secondary response to glucocorticoids *in vivo*. *Proc. Natl. Acad. Sci. USA* **87:** 2564-2568.

Hess, P., Meenakshi, T., Chan, G.C-K., Carlstedt-Duke, J., Gustafsson, J-A., and Payvar, F. (1990). Purified glucocorticoid receptors bind selectively *in vitro* to a cloned DNA fragment that mediates a delayed secondary response to glucocorticoids *in vivo*. *Proc. Natl. Acad. Sci. USA* **87:** 2564-2568.

Hoheisel, J.D., Maier, E., Mott, R., McCarthy, L., Grigoriev, A.V., Schalkwyk, L.C., Nizetic, D., Francis, F., and Lehrach, H. (1993). High-resolution cosmid and P1 maps spanning the 14-Mb genome of the fission yeast *S. pombe*. *Cell* **73:** 109-120.

Holzfeind, P., and Redl, B. (1994). Structural organization of the gene encoding the human lipocalin tear prealbumin and synthesis of the recombinant protein in *Escherichia coli*. *Gene* **139:** 177-183.

Hood, L., Campbell, J.H., and Elgin, S.C.R. (1975). The organization, expression, and evolution of antibody genes and other multigene families. *Ann. Rev. Genet.* **9:** 305-353.

Houseal, T.W., Dackowski, W.R., Landes, G.M., and Klinger, K.W. (1994). High resolution mapping of overlapping cosmids by fluorescence *in situ* hybridization. *Cytometry* **15:** 193-198.

Husman, B., Norstedt, G., Mode, A., and Gustafsson, J-A. (1985). The mode of growth hormone administration is of major importance for the excretion of the major male rat urinary proteins. *Mol. Cell. Endocrin.* **40:** 205-210.

Igarashi, M., Nagata, A., Toh, H., Urade, H., and Hayaishi, O. (1992). Structural organisation of the gene for prostaglandin synthase in the rat brain. *Proc. Natl. Acad. Sci. USA* **89:** 5376-5380.

Ioannou, P.A., Amemiya, C.T., Garnes, J., Kroisel, P.M., Shizuya, H., Chen, C., Batzer, M.A., and de Jong, P.J. (1994). A new bacteriophage P1-derived vector for the propagation of large human DNA fragments. *Nature Genet.* **6:** 84-89.

Irwin, J.F., Lane, S.E., and Neuhaus, O.W. (1971). Synergistic effect of glucocorticoids and androgens on the biosynthesis of a sex-dependent protein in the male rat. *Biochim. Biophys. Acta.* **252:** 328-334.

Jacob, H.J., Brown, D.M., Bunker, R.K., Daly, M.J., Dzau, V.J., Goodman, A., Koike, G., Kren, V., Kurtz, T., Lemmark, A., Levan, G., Mao, Y., Pettersson, A., Pravenec, M., Simon, J.S., Szpirer, C., Szpirer, J., Trolliet, M.R., Winer, E.S., and Lander, E.S. (1995). A genetic linkage map of the laboratory rat, *Rattus norvegicus*. *Nature Genetics* **9:** 63-69.

Jeffreys, A.J., Wilson, V., and Thein, S.L. (1985). Hypervariable "minisatellite" regions in human DNA. *Nature* **314:** 67-73.

Johnson, D., Harrison, S., Pineda, N., Heinlein, C., Al-Shawi, R., and Bishop, J.O. (1995). Localization of the response elements of a gene induced by intermittent growth hormone stimulation. *J. Mol. Endocrin.* **14:** 35-49.

Julkunen, M., Seppala, M., and Janne, O.A. (1988). Complete amino acid sequence of human placental protein 14: A progesterone-regulated uterine protein homologous to beta-lactoglobulins. *Proc. Natl. Acad. Sci. USA* **85:** 8845-8849.

Kanai, M., Raz, A., and Goodman, D.S. (1968). Retinol-binding protein: The transport for vitamin A in human plasma. *J. Clin. Invest.* **47:** 2025-2044.

Kaumeyer, J.F., Polazzi, J.O., and Kotic, M.P. (1986). The mRNA for a proteinase inhibitor related to the HI-30 domain of inter-alpha-trypsin inhibitor also encodes alpha-1-microglobulin (protein HC). *Nucl. Acids Res.* **14:** 7839-7850.

Keen, J.N., Caceres, I., Eliopoulos, E.E., Zagalsky, P.F., and Findlay, J.B.C. (1991b). Complete sequence and model for the C_1 subuniy of the carotenoprotein, crustacyanin, and model for the dimer beta-

crustacyanin, formed from the C_1 and A_2 subunits with astaxanthin. *Eur. J. Biochem.* **202:** 31-40.

Keen, J.N., Caceres, I., Eliopoulos, E.E., Zagalsky, P.F., and Findlay, J.B.C. (1991a). Complete sequenceand model for the A_2 subunit of the carotenoid pigment complex, crustacyanin. *Eur. J. Biochem.* **197:** 407-417.

Kit, S., Dubbs, D.R., Piekarski, L.J., and Hsu, T.C. (1963). Deletion of thymidine kinase activity from L cells resistant to bromodeoxyuridine. *Exp. Cell Res.* **31:** 297-312.

Kjeldsen, L., Bainton, D.F., Sengelov, H., and Borregaard, N. (1994). Identification of neutrophil gelatinase-associated lipocalin as a novel matrix protein of specific granules in human neutrophils. *Blood* **83:** 799-807.

Kjeldsen, L., Johnsen, A.H., Sengelov, H., and Borregaard, N. (1993). Isolation and primary structure of NGAL, a novel protein associated with human neutrophil gelatinase. *J. Biol. Chem.* **268:** 10425-10432.

Knopf, J.L., Gallagher, J.F., and Held, W.A. (1983). Differential, multihormonal regulation of the mouse major urinary protein gene family in the liver. *Mol. Cell. Biol.* **3:** 2232-2240.

Kock, K., Ahlers, C., and Schmale, H. (1994). Structural organization of the genes for rat von Ebner's gland proteins 1 and 2 reveals their close relationship to lipocalins. *Eur. J. Biochem.* **221:** 905-916.

Kolata, G. (1987). What babies know, and noises parents make. *Science* **237:** 726.

Korenberg, J.R., and Rykowski, M.C. (1988). Human genome organization: Alu, lines, and the molecular structure of metaphase chromosome bands. *Cell* **53:** 391-400.

Korf, G.M., and Stumph, W.E. (1986). Chicken U2 and U1 RNA genes are found in very different genomic environments but have similar promoter structures. *Biochemistry* **25:** 2041-2047.

Kremer, H., and Hennig, W. (1990). Isolation and characterization of a *Drosophila hydei* histone DNA repeat unit. *Nucl. Acids Res.* **18:** 1573-1580.

Kretz, P.L., Reid, C.H., Greener, A., and Short, J.M. (1989). Effect of lambda packaging extract mcr restriction activity on DNA cloning. *Nucl. Acids Res.* **17:** 5409.

Krumlauf, R. (1992). Evolution of the vertebrate Hox homeobox genes. *BioEssays* **14:** 245-252.

Kumar, M., Roy, A.K., and Axelrod, A.E. (1969). Androgenic induction of α_{2u} globulin in the rat: Requirement of an intact pituitary. *Nature* **223:** 399-400.

Kurtz, D.T. (1981). Rat α_{2u} globulin is encoded by a multigene family. *J. Mol. Appl. Genet.* **1:** 29-38.

Kurtz, D.T. (1981b). Hormonal inducibility of rat α_{2u} globulin genes in transfected mouse cells. *Nature* **291:** 629-631.

Kurtz, D.T., and Feigelson, P. (1978). Multihormonal control of the messenger RNA for the hepatic protein α_{2u} globulin. *In* "Biochemical Actions of Hormones V". pp. 433-455, Academic Press, Inc., New York.

Kurtz, D.T., Chan, K-M., and Feigelson, P. (1978a). Translational control of hepatic α_{2u} globulin synthesis by growth hormone. *Cell* **15:** 743-750.

Kurtz, D.T., Chan, K-M., and Feigelson, P. (1978b). Glucocorticoid induction of hepatic α_{2u} globulin synthesis and messenger RNA level in castrated male rats *in vivo*. *J. Biol. Chem.* **253:** 7886-7890.

Kurtz, D.T., McCullough, L., Bishop, D., and Manos, M.M. (1982). DNA sequences required for hormonal induction of rat α_{2u} globulin genes. *Cold Spring Harbor Symp. Quant. Biol.* **47:** 985-988.

Kurtz, D.T., Sippel, A.E., Ansah-Yiadom, R., and Feigelson, P. (1976). Effects of sex hormones on the level of the messenger RNA for the rat hepatic protein α 2u globulin. *J. Biol. Chem.* **251:** 3594-3598.

Laird, C.D., McConaughy, B.L., and McCarthy, B.J. (1969). Rate of fixation of nucleotide substitutions in evolution. *Nature* **224:** 149-154.

Lalley, P.A., Peters, J., Doolittle, D.P., Hillyard, A.L., and Searle, A.G. (1991). Report on the comparative committee for human, mouse and other rodents. *Cytogenet. Cell Genet.* **58:** 1152-1189.

Laperche, Y., Lynch, K.R., Dolan, K.P., and Feigelson, P. (1983). Tissue-specific control of α_{2u} globulin gene expression: Constitutive synthesis in the submaxillary gland. *Cell* **32:** 453-460.

Larin, Z., Monaco, A.P., and Lehrach, H. (1991). Yeast artificial chromosome libraries containing large inserts from mouse and human DNA. *Proc. Natl. Acad. Sci. USA* **88:** 4123

Lawrence, J.B., Singer, R.H., and McNeil, J.A. (1990). Interphase and metaphase resolution of different distances within the human dystrophin gene. *Science* **249:** 928-932.

Lehman-McKeeman, L.D. (1993). Male rat-specific light hydrocarbon nephropathy. *In* "Toxicology of the Kidney". pp. 477-494, Raven Press, New York.

Lehman-McKeeman, L.D., Caudill, D., Takigiku, R., Schneider, R.E., and Young, J.A. (1990b). Comparative disposition of d-limonene in rats and mice: Relevance to male-rat-specific nephrotoxicity. *Toxicol. Lett.* **53:** 193-195.

Lehman-McKeeman, L.D., Rivera-Torres, M.I., and Caudill, D. (1990a). Lysosomal degradation of α_{2u}-globulin and α_{2u}-globulin-xenobiotic conjugates. *Toxicol. Appl. Pharmacol.* **103:** 539-548.

Lehman-McKeeman, L.D.,Rodriguez, P.A., Caudill, D., Fey, M.L., Eddy, C.L., and Asquith, T.N. (1991). Hyaline droplet nephropathy resulting from exposure to 3,5,5-trimethylhexanoyloxybenzene sulfonate. *Toxicol. Appl. Pharmacol.* **107:** 429-438.

Lehman-McKeeman, L.D.,Rodriguez, P.A., Takigiku, R., Caudill, D., and Fey, M.L. (1989). D-Limonene-induced male rat-specific nephrotoxicity: Evaluation of the association between D-Limonene and α_{2u}-globulin. *Toxicol. Appl. Pharmacol.* **99:** 250-259.

Lengauer, C., Speicher, M.R., Popp, S., Jauch, A., Taniwaki, M., Nagaraja, R., Riethman, H.C., Donis-Keller, H., D'Urso, M., Schlessinger, D., and Cremer, T. (1993). Chromosomal bar codes produced by multicolor fluorescence *in situ* hybridization with mutiple YAC clones and whole chromosome painting probes. *Hum. Mol. Genet.* **2:** 505-512.

Levan, G., Szpirer, J., Szpirer, C., Klinga, K., Hanson, C., and Islam, M.Q. (1991). The gene map of the Norway rat (*Rattus norvegicus*) and comparative mapping with mouse and man. *Genomics* **10:** 699-718.

Liao, Y.C., Taylor, J.M., vannice, J.L., Clawson, G.A., and Smuckler, E.A. (1985). Structure of the rat alpha-1-acid glycoprotein gene. *Mol. Cell. Biol.* **5:** 3634-3639.

Lica, L.M., Narayanswami, S., and Hamkalo, B.A. (1986). Mouse satellite DNA, centromere structure, and sister chromatid pairing. *J. Cell Biol.* **103:** 1145-1151.

Lichter, P., Tang, C.C., Call, K., Hermanson, G., Evans, G., Housman, D, and Ward, D.C. (1990). High-resolution mapping of human chromosome 11 by *in situ* hybridization with cosmid clones. *Science* **247:** 64-69.

Liebhaber, S.A., Goosens, M.A., and Kan, Y.W. (1981). Homology and concerted evolution of the alpha 1 and alpha 2 loci of human alpha-globin. *Nature* **290:** 26-29.

Lindgren, V., Ares, M., Weiner, A.M., and Francke, U. (1985). Human genes for U2 small nuclear RNA map to a major adenovirus 12 modification site on chromosome 17. *Nature* **314:** 115-116.

Lindsey, J.R. (1979). Historical foundations in the laboratory rat. *In* "The Laboratory Rat". pp. 1-36, Academic Press Inc., New York.

Little, R.D., Porta, G., Carle, G.F., Schlessinger, D., and D'Urso, M. (1989). Yeast artificial chromosomes with 200- to 800- kilobase inserts of human DNA containing HLA, V kappa, 5S, and Xq24-Xq28 sequences. *Proc. Natl. Acad. Sci. USA* **86:** 1598-1602.

Lock, E.A., Charbonneau, M., Strasser, J., Swenberg, J.A., and Bus, J.S. (1987). 2,2,4-Trimethylpentane-induced nephrotoxicity. II. The reversible binding of TMP metabolite to a renal protein fraction containing α_{2u}-globulin. *Toxicol. Appl. Pharmacol.* **91:** 182-192.

Loeb, L.A. (1989). Endogenous carcinogenesis: Molecular oncology into the twenty-first century-Presidential Address. *Cancer Res.* **49:** 5489-5496.

Logdberg, L., and Akerstrom, B. (1981). Immunosuppressive properties of alpha-1-microglobulin. *Scand. J. Immun.* **13:** 383-390.

Long, E.O., and Dawid, I.B. (1980). Repeated genes in eukaryotes. *Ann. Rev. Biochem.* **49:**727-764.

Lozovskaya, E.R., Petrov, D.A., and Hartl, D.L. (1993). A combined molecular and cytogenetic approach to genome evolution in

Drosophila using large fragment DNA cloning. *Chromosoma* **102:** 253-266.

Lund, E., Dahlberg, J.E., and Forbes, D.E. (1984). The two embryonic U1 small nuclear RNAs of *Xenopus laevis* are encoded by a major family of tandemly repeated genes. *Mol. Cell. Biol.* **4:** 2580-2586.

Lundin, L.G. (1993). Evolution of the vertebrate genome as reflected in paralogous chromosomal regions in man and in the house mouse. *Genomics* **16:** 1-19.

Lynch, K.R., Dolan, K.P., Nakhasi, H.L, Unterman, R., and Feigelson, P. (1982). The role of growth hormone in α_{2u} globulin synthesis: A reexamination. *Cell* **28:** 185-189.

MacInnes, J.I., Nozik, E.S., Kurtz, D.T. (1986). Tissue-specific expression of the rat alpha$_{2u}$ globulin gene family. *Mol. Cell. Biol.* **6:** 3563-3567.

MacIntyre, R.J. (1994). Molecular Evolution: Codes, clocks, genes and genomes. *Bioessays* **16:** 699-703.

Macrides, F., Singer, A.G., Clancy, A.N., Goldman, B.D., and Agosta, W.C. (1984). Male hamster investigatory and copulatory responses to vaginal discharge: An attempt to impart sexual significance to an arbitrary chemosensory stimulus. *Physiol. and Behav.* **33:** 627-632.

Maeda, N., and Smithies, O. (1986). The evolution of multigene families: Human haptoglobin genes. *Ann. Rev. Genet.* **20:** 81-108.

Mancini, M., Majumdar, D., Chatterjee, B., and Roy, A.K. (1989). α_{2u}-globulin in modified sebaceous glands with pheromonal functions: Localization of the protein and its mRNA in preputial, meibomian, and perianal glands. *J. Histochem. Cytochem.* **37:** 149-157.

Manser, T., and Gesteland, R.F. (1982). Human U1 loci: Genes for human U1 RNA have dramatically similar genomic environments. *Cell* **29:** 257-264.

Marzluff, W.E., and Graves, R.A. (1984). Organization and expression of mouse histone genes. *In* "Histone Genes. Structure, Organization and Regulation". pp. 281-315, John Wiley & Sons, New York.

Matera, A.G., Weiner, A.M., and Schmid, C.W. (1990). Structure and evolution of the U2 small nuclear RNA multigene family in primates:

gene amplification under natural selection? *Mol. Cell. Biol.* **10:** 5876-5882.

Mathias, S.L., Scott, A.F., Kazazian, H.J., Boeke, J.D., and Gabriel, A. (1991). Reverse transcriptase encoded by a human transposable element. *Science* **254:** 1808-1810.

Mattaj, I.W., and Zeller, R. (1983). Xenopus laevis U2 snRNA genes: Tandemly repeated transcription units sharing 5' and 3' homology with other RNA polymerase II transcribed genes. *EMBO J.* **2:** 1883-1891.

Maxson, R., Cohn, R., Kedes, L., and Mohun, T. (1983). Expression and organization of histone genes. *Ann. Rev. Genet.* **17:** 239-277.

McClintock, M.K. (1971). Menstrual synchrony and suppression. *Nature* **229:** 244-245.

McCormack, W.T., and Thompson, C.B. (1990). Chicken Ig(L) variable region gene conversions display pseudogene donor preference and 5' to 3' polarity. *Genes Dev.* **4:** 548-558.

McCormick, M.K., Shero, J.H., Cheung, M.C., Kan, Y.W., Hieter, P.A., and Antonarakis, S.E. (1989). Construction of human chromosome 21 specific yeast artificial chromosome. *Proc. Natl. Acad. Sci. USA* **86:** 9991-9995.

McCutchan, T.F. (1986). The ribosomal genes of *Plasmodium*. *Int. Rev. Cytol.* **99:** 295-309.

Meeks-Wagner, D., and Hartwell, L.H. (1986). Normal stoichiometry of histone dimer sets is necessary for high fidelity of mitotic chromosome transmission. *Cell* **44:** 43-52.

Melrose, D.R., Reed, H.C., and Patterson, R.L. (1971). Androgen steroids associated with boar odour as an aid to the detection of oestrus in pig artificial insemination. *Brit. Vet. J.* **127:** 497-502.

Mendez, E., Fernandez-Luna, J.L., Grubb, A., and Leyva-Cobain, F. (1986). Human protein HC and its IgA complex are inhibitors of neutrophil chemotaxis. *Proc. Natl. Acad. Sci. USA* **83:** 1472-1475.

Menzies, R.A., Heth, G., Ikan, R., Weinstein, V., and Nevo, E. (1992). Sexual pheromones in lipids and other fractions from urine of the male mole rat, *Spalax ehrenbergi*. *Phsiol. Behav.* **52:** 741-747.

Meyne, J., Ratliff, R.L., and Moyzis, R.K. (1989). Conservation of the human telomere sequence $TTAGGG_n$ among vertebrates. *Proc. Natl. Acad. Sci USA* **86:** 7049-7053.

Miller, J. (1972). *In* "Experiments in Molecular Genetics". Cold Springs Harbor Press, Cold Spring Harbor.

Miller, S.J., Aly, R., Shinefeld, H.R., and Elias, P.M. (1988). *In vitro* and *in vivo* antistaphylococcal activity of human stratum corneum lipids. *Arch. Dermatol.* **124:** 209-215.

Mirsky, A.E. and Ris, H. (1948). Isolated chromosomes. *J. Gen. Phsiol.* **31:** 1-6.

Mode, A., Gustafsson, J.A., Jansson, J.O., Eden, S., and Isaksson, O. (1982). Association between plasma level of growth hormone and sex differentiation of hepatic steroid metabolism in the rat. Endocrinology 111: 1692-1697.

Monaco, A.P., and Larin, Z. (1994). YACs, BACs, PACs and MACs: Artificial chromosomes as research tools. *TibTech.* **12:** 280-286.

Monaco, H.L., Zanotti, G., Spadon, P., Bolognesi, M., Sawyer, L., and Eliopoulos, E.E. (1987). Crystal structure of the tigonal form of bovine beta-lactoglobulin and of its complex with retinol at 2.5 A resolution. *J. Mol. Biol.* **197:** 695-706.

Morzycka-Wroblewska, E., Selker, E.U., Stevens, J.N., and Metzenberg, R.L. (1985). Concerted evolution of dispersed *Neurospora crassa* 5S rRNA genes: Pattern of sequence conservation between allelic and non-allelic genes. *Mol. Cell. Biol.* **5:** 46-51.

Murty, C.V.R., Sarkar, F.H., Mancini, M.A., and Roy, A.K. (1987). Sex-independent synthesis of α_{2u}-globulin and its mRNA in the rat preputial gland. *Endocrinology* **121:** 1000-1005.

Nagata, A., Suzuki, Y., Igarashi, M., Eguchi, N., Toh, H., Urade, Y., and Hayaishi, O. (1991). Human brain prostaglandin D synthase has been evolutionarily differentiated from lipophilic-ligand carrier proteins. *Proc. Natl. Acad. Sci. USA* **88:** 4020-4024.

Nakamura, Y., Leppert, M., O'Connell, P., Wolff, R., Holm, T., Culver, M., Martin, C., Fujimoto, E., Hoff, M., and Kumlin, E. (1987). Variable number of tandem repeat (VNTR) markers for human gene mapping. *Science* **235:** 1616-1622.

Nesbitt, M.N. (1974). Evolutionary relationships between rat and mouse chromosomes. *Chromosoma* **46:** 217-224.

Neuhaus, O.W., Flory, W., Biswas, N, and Hollerman, C.E. (1981). Urinary excretion of α_{2u}-globulin and albumin by adult male rats following treatment with nephrotoxic agents. *Nephron* **28:** 133-140.

Nevo, E., and Heth, G. (1976). Assortive mating between chromosome forms of the mole rat, *Spalax ehrenbergi*. *Experientia* **32:** 1509-1511.

Nevo, E., Bodmer, M., and Heth, G. (1976). Olfactory discrimination as an isolating mechanism in speciating mole rats. *Experientia* **32:** 1511-1513.

Newcomer, M.E. (1993). Structure of the epididymal retinoic acid binding protein at 2.1 A resolution. *Structure* **1:** 7-18.

Newcomer, M.E., Jones, T.A., Aqvist, J., Sundelin, J., Eriksson, U., Rask, L., and Peterson, P.A. (1984). The three-dimensional structure of retinol-binding protein. *EMBO J.* **3:** 1451-1454.

Nikaido, H., Yamada, J., and Kondo, Y. (1982). Male urinary protein-1 (MUP-1) in the rat. *J. Hered.* **73:** 119-122.

Nojima, H., and Kornberg, R.D. (1983). Genes and pseudogenes for mouse U1 and U2 small nuclear RNAs. *J. Biol. Chem.* **258:** 8151-8155.

North, A.C.T. (1991). Structural homology in ligand-specific transport proteins. *Biochem. Soc. Symp.* **57:** 35-48.

Nur, I., Pascale, E., and Furano, A.V. (1988). The left end of rat L1 (L1Rn, long interspersed repeated) DNA which is a CpG island can function as apromoter. *Nucl. Acids Res.* **16:** 9233-9251.

O'Brien, S.J., and Graves, J.A.M. (1991). Report of the committee on comparative gene mapping. *Cytogenet. Cell Genet.* **58:** 1124-1151.

Oakey, R.J., Caron, M.G., Lefkowitz, R.J., and Seldin, M.F. (1991). Genomic organization of adrenergic and serotonin receptors in the mouse: Linkage mapping of sequence-related genes provides a method for examining mammalian chromosome evolution. *Genomics* **10:** 338-344.

Ohno, S. (1970). Evolution by Gene Duplication. Springer-Verlag, New York.

Ohta, T. (1988). Multigene and supergene families. *In* "Oxford Surveys in Evolutionary Biology". pp. 41-65, Oxford University Press, Oxford.

Ohta, T. (1991). Multigene families and their implications for evolutionary theory. *In* "New Aspects of the Genetics of Molecular Evolution". pp. 15-25, Springer-Verlag, Berlin.

Olson, M.J., Garg, B.D., Murty, C.V.R., and Roy, A.K. (1987). Accumulation of α_{2u}-globulin in the renal proximal tubules of male rats exposed to unleaded gasoline. *Toxicol. Appl. Pharmacol.* **90:** 43-51.

Papiz, M.Z., Sawyer, L., Eliopoulos, E.E., North, A.C., Findlay, J.B., Sivaprasadarao, R., Jones, T.A., Newcomer, M.E., and Kraulis, P.J. (1986). The structure of beta-lactoglobulin and its similarity to plasma retinol-binding protein. *Nature* **324:** 383-385.

Paquereau, L., Vilarem, M.J., Rossi, V., Rouayrenc, J.F., and Le Cam, A. (1992). Regulation of two rat serine-prorease inhibitor gene promoters by somatotropin and glucocorticoids. Study with intact hepatocytes and cell-free systems. *Eur. J. Biochem.* **209:** 1053-1061.

Pardue, M.L. and Gall, J.G. (1970). Chromosomal localisation of mouse satellite DNA. *Proc. Natl. Acad. Sci. USA* **168:** 1356-1358.

Parkes, A.S., and Bruce, H.M. (1962). Pregnancy-block in female mice placed in boxes soiled by males. *J. Reprod. Fertil.* **4:** 303-308.

Parra, I., and Windle, B. (1993). High resolution visual mapping of stretched DNA by fluorescent hybridization. *Nature Genet.* **5:** 17-21.

Pascale, E., Liu, C., Valle, E., Usdin, K., and Furano, A.V. (1993). The evolution of long interspersed repeated DNA (L1, LINE 1) as revealed by the analysis of an ancient rodent L1 DNA family. *J. Mol. Evol.* **36:** 9-20.

Peavy, D.E., Taylor, J.M., and Jefferson, L.S. (1981). Alterations in albumin secretion and total protein synthesis in livers of thyroidectomized rats. *Biochem. J.* **198:** 289-299.

Pelosi, P. (1994). Odorant-binding proteins. *Crit. Rev. Biochem. Mol. Biol.* **29:** 199-228.

Pervaiz, S., and Brew, K. (1985). Homology of β-lactoglobulin, serum retinol-binding protein and protein HC. *Science* **228:** 335-337.

Pes, D., and Pelosi, P. (1995). Odorant-binding proteins of the mouse. *Comp. Biochem. Physiol.* **112B:** 471-479.

Pes, D., Dal Monte, M., Ganni, M., and Pelosi, P. (1992). Isolation of two odorant-binding proteins from mouse nasal tissue. *Comp. Biochem. Physiol.* **103B:** 1011-1017.

Pevsner, J., and Snyder, S.H. (1990). Odorant-binding protein: Odorant transport function in the vertebrate nasal epithelium. *Chem. Senses* **15:** 217-222.

Pevsner, J., Reed, R.R., Feinstein, P.G., and Snyder, S.H. (1988). Molecular cloning of odorant-binding protein: Member of a ligand carrier family. *Science* **241:** 336-339.

Pevsner, J., Sklar, P.B., and Snyder, S.H. (1986). Odorant-binding protein: Localization to nasal glands and secretions. *Proc. Natl. Acad. Sci. USA* **83:** 4942-4946.

Pevsner, J., Trifiletti, R.R., Strittmatter, S.M., and Snyder, S.H. (1985). Isolation and characterization of an olfactory receptor protein for odorant pyrazines. *Proc. Natl. Acad. Sci. USA* **82:** 3050-3054.

Pierce, J.C., Sternberg, N., and Sauer, B. (1992). A mouse genomic library in the bacteriophage-P1 cloning system- organization and characterization. *Mamm. Genome* **3:** 550-558.

Pinkel, D., Straume, T., and Gray, J.W. (1986). Cytogenetic analysis using quantitative, high-sensitivity, fluorescence hybridization. *Proc. Natl. Acad. Sci. USA* **83:** 2934-2938.

Pizzuti, A., Novelli, G., Mari, A., Ratti, A., Colosimo, A., Amati, F., Penso, D., Sangiuolo, F., Calabrese, G., Palka, G., Silani, V., Gennarelli, M., Mingarelli, R., Scarlato, G., Scambler, P., and Dallapiccola, B. (1996). Human homologue sequences to the *Drosophila* dishevelled segment-polarity gene are deleted in the DiGeorge Syndrome. *Am. J. Hum. Genet.* **58:** 722-729.

Preti, G., Cutler, W.B., Garcia, G.R., Huggins, G.R., and Lawley, H.J. (1986). Human axillary secretions influence women's menstrual cycles: The role of donor extract on females. *Horm. Behav.* **20:** 474-482.

Radic, M.L., Lundgren, K., and Hamkalo, B.A. (1987). Curvature of mouse satellite DNA and condensation of heterochromatin. *Cell* **50:** 1101-1108.

Raming, K., Krieger, J., Strotmann, J., Boekhoff, I., Kubic, S., Baumstark, C., and Breer, H. (1993). Cloning and expression of odorant receptors. *Nature* **361:** 353-356.

Read, N.G., Astbury, P.J., Morgan, R.J.I., Parsons, D.N., and Port, C.J. (1988). Induction and exacerbation of hyaline droplet formation in the proximal tubular cells of the kidneys of male rats receiving a variety of pharmacological agents. *Toxicology* **52:** 81-101.

Redl, B., Holzfeind, P., and Lottspeich, F. (1992). cDNA cloning and sequencing reveals human tear prealbumin to be a member of the lipophilic-ligand carrier protein superfamily. *J. Biol. Chem.* **267:** 20282-20287.

Ried, T., Baldini, A., Rand, T., and Ward, D.C. (1992). Simultaneous visualization of seven different DNA probes by *in situ* hybridization using combinatorial fluorescence and digital imaging microscopy. *Proc. Natl. Acad. Sci. USA* **89:** 1388-1392.

Riley, C.T., Barbeau, B.K., Keim, P.M., Kezdy, F.J., Heinrikson, R.L., and Law, J.H. (1984). The covalent protein structure of insecticyanin, a blue biliprotein from the hemolymph of the tobacco hornworm, *Manduca sexta* L. *J. Biol. Chem.* **259:** 13159-13165.

Riley, J.H., Ogilvie, D., and Anand, R. (1992). Construction, characterization and screening of YAC libraries. *In* "Techniques for the Analysis of Complex Genomes". pp. 59-79, Academic Press Inc, New York.

Roberts, S.B., Sanicola, M., Emmons, S.W., and Childs, G. (1987). Molecular characterization of the histone gene family of Caenorhabditis elegans. *J. Mol. Biol.* **196:** 27-38.

Roberts, S.B., Weisser, K.E., and Childs, G. (1984). Sequence comparisons of non-allelic late histone genes and their early stage counterparts. Evidence for gene conversion with the sea urchin late stage gene family. *J. Mol. Biol.* **174:** 647-662.

Robinson, C. (1992). The genome race is on the road. *TibTech* **10:** 1-5.

Rogers, J.H. (1985). The origin and evolution of retroposons. *Int. Rev. Cytol.* **93:** 187-279.

Roy, A.K. (1973). Androgen-dependent synthesis of α_{2u} globulin in the rat. Role of the pituitary gland. *J. Endocrinol.* **56:** 295-301.

Roy, A.K. (1977). Early events in the steroidal regulation of α_{2u} globulin in rat liver. *Eur. J. Biochem.* **73:** 537-543.

Roy, A.K. (1979). Hormonal regulation of α_{2u} globulin in rat liver. *In* "Biochemical Actions of Hormones VI". pp. 481-516, Academic Press Inc., New York.

Roy, A.K. (1979). Hormonal regulation of α_{2u} globulin in rat liver. *In* "Biochemical Actions of Hormones VI". pp. 481-516, Academic Press, Inc., New York.

Roy, A.K. and Neuhaus, O.W. (1966a). Identification of rat urinary protein by zone and immunoelectrophoresis. *Proc. Soc. Exp. Biol. Med.* **121:** 894-899.

Roy, A.K. and Neuhaus, O.W. (1966b). Proof of the hepatic synthesis of a sex-dependent protein in the rat. *Biochim. Biophys. Acta* **127:** 82-87.

Roy, A.K., and Dowbenko, D.J. (1977). Role of growth hormone in the multihormonal regulation of messenger RNA for α_{2u} globulin in the liver of hypophysectomized rats. *Biochemistry* **16:** 3918-3922.

Roy, A.K., and Leonard, S. (1973). Androgen-dependent synthesis of α2u globulin in diabetic rats: The role of insulin. *J. Endocrinol.* **57:** 327-328.

Roy, A.K., and Neuhaus, O.W. (1967). Androgenic control of a sex-dependent protein in the rat. *Nature* **214:** 618-620.

Roy, A.K., and Raber, D.L. (1972). Immunofluorescent localization of α_{2u} globulin in the hepatic and renal tissues of rat. *J. Histochem. Cytochem.* **20:** 89-96.

Roy, A.K., Chatterjee, B., Demyan, W.F., Nath, T.S., and Motwani, N.M. (1982). Pretranslational regulation of α_{2u} globulin in rat liver by growth hormone. *J. Biol. Chem.* **257:** 7834-7838.

Roy, A.K., McMinn, D., and Biswas, N.M. (1975). Estrogenic inhibition of the hepatic synthesis of α_{2u}-globulin in the rat. *Endocrinology* **97:** 1501-1508.

Roy, A.K., Milin, B.S., and McMinn, D.M. (1974). Androgen receptor in rat liver: Hormonal and developmental regulation of the cytoplasmic and its correlation with the androgen dependent synthesis of α_{2u} globulin. *Biochim. Biophys. Acta.* **354:** 213-232.

Roy, A.K., Neuhaus, O.W., and Harmison, C.R. (1966). Preparation and characterization of a sex-dependent rat urinary protein. *Biochim. Biophys. Acta.* **127:** 72-81.

Royle, N.J., Clarkson, R.E., Wong, Z., and Jeffreys, A.J. (1988). Clustering of hypervariable minisatellites in the proterminal regions of human autosomes. *Genomics* **3:** 352-360.

Russell, M.J., Switz, G.M., and Thompson, K. (1980). Olfactory influences on the human menstrual cycle. *Pharmacol. Biochem. Behav.* **13:** 737-738.

Salier, J.P., Simon, D., Rouet, P., Raguenez, G., Muscatelli, F., Gebhard, W., Guenet, J.L, and Mattei, M.G. (1992). Homologous chromosomal locations of the four genes for inter-α-inhibitor and pre-α-inhibitor family in human and mouse: assignment of the ancestral gene for the lipocalin superfamily. *Genomics* **14:** 83-88.

Sansom, C.E., North, A.C.T., and Sawyer, L. (1994). Structural analysis and classification of lipocalins and related proteins using a profile-search method. *Biochim. Biophys. Acta* **1208:** 247-255.

Sarkar, P., and Feigelson, P. (1989). Developmentally regulated male-specific transfactor(s) enable in vitro transcription of acloned alpha-2u-globulin gene. *Mol. Endocrin.* **3:** 342-348.

Schalkwyk, L.C., Francis, F., and Lehrach, H. (1995). Techniques in mammalian genome mapping. *Curr. Opin. Biotech.* **6:** 37-43.

Schmeckpeper, B.J., Scott, A.F., and Smith, K.D. (1984). Transcripts homologous to a long repeated DNA element in the human genome. *J. Biol. Chem.* **259:** 1218-1225.

Scnmid, C., and Maraia, R. (1992). Transcriptional regulation and transpositional selection of active SINE sequences. *Curr. Opin. Genet. Dev.* **2:** 874-882.

Schubert, D. and LaCorbiere, M. (1985). Isolation of an adhesion-mediating protein from chick neural retina adhesions. *J. Cell Biol.* **101:** 1071-1077.

Schubert, D., LaCorbiere, M., and Esch, F. (1986). *J. Cell Biol.* **102:** 2295-22301.

Scott, G.K., Hayes, P.H., Fletcher, G.L., and Davies, P.L. (1988). Wolffish antifreeze proteingenes are primarily organized as tandem repeats that each contain two genes in inverted orientation. *Mol. Cell. Biol.* **8:** 3670-3675.

Selker, E.U., Yanofsky, C., Driftmier, K., Metzenberg, R.L., Alzner-DeWeerd, B., and Rajbhandary, U.L. (1981). Dispersed 5S RNA genes in *N. crassa*: Structure, expression and evolution. *Cell* **24:** 819-828.

Shahan, K., Denaro, M., Gilmartin, M., Shi, Y., and Derman, E. (1987b). Expression of six mouse major urinary protein genes in the mammary, parotid, sublingual, submaxillary, and lachrymal glands and in the liver. *Mol. Cell. Biol.* **7:** 1947-1954.

Shahan, K., Gilmartin, M., and Derman, E. (1987a). Nucleotide sequences of liver, lachrymal, and submaxillary gland mouse major urinary protein mRNAs: Mosaic structure and construction of panels of gene specific synthetic oligonucleotide probes. *Mol. Cell. Biol.* **7:** 1938-1946.

Shaw, P.H., Held, W.A., and Hastie, N.D. (1983). The gene family for major urinary proteins: Expression in several secretory tissues of the mouse. *Cell* **32:** 755-761.

Shepherd, N.S., Pfrogner, B.D., Coulby, J.N., Ackerman, S.L., Vaidyanathan, G., Sauer, R.H., Balkenhol, T.C., and Sternberg, N. (1994). Preparation and screening of an arrayed human genomic library generated with the P1 cloning system. *Proc. Natl. Acad. Sci. USA* **91:** 2629-2633.

Shi, Y., Son, H.J., Shahan, K., Rodriguez, M., Costantini, F., and Derman, E. (1989). Silent genes in the mouse major urinary protein gene family. *Proc. Natl. Acad. Sci. USA* **86:** 4584-4588.

Shizuya, H., Birren, B., Kim, U.J., Mancino, V., Slepak, T., Tachiiri, Y., and Simon, M. (1992). Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in Escherichia coli using an F-factor-based vector. *Proc. Natl. Acad. Sci. USA* **89:** 8794-8797.

Short, B.G., Burnett, V.L., and Swenberg, J.A. (1989). Elevated proliferation of proximal tubule cells and localization of accumulated α_{2u}-globulin in F344 rats during chronic exposure to unleaded gasoline or 2,2,4-trimethylpentane. *Toxicol. Appl. Pharmacol.* **101:** 414-431.

Sierra, F., Tamone, F., Mueller, C.R., and Schibler, U. (1990). Differential in vitro transcription from the promoter of a rat alpha-2u-globulin gene in liver and spleen nuclear extracts. *Mol. Biol. Med.* **7:** 131-146.

Singer, A.G., and Macrides, F. (1990). Aphrodisin: Pheromone or transducer? *Chem. Senses* **15:** 199-203.

Singer, A.G., Macrides, F., Clancy, A.N., and Agosta, W.C. (1986). Purification and analysis of a proteinaceous aphrodisiac pheromone from hamster vaginal discharge. *J. Biol. Chem.* **261:** 13323-13326.

Singer, M., and Berg, P. (1991). Genes and Genomes: A changing perspective. University Science Books, Mill Valley.

Singer, M.F., Thayer, R.E., Grimaldi, G., Lerman, M.J., and Fanning, T.G. (1983). Homology between the *Kpn*I primate and *Bam* HI (MIF-1) rodent families of long interspersed, repeated sequences. *Nucl. Acids Res.* **11:** 5739-5745.

Sinnett, D., Richer, C., deragon, J-M., and Labuda, D. (1991). *Alu* RNA secondary structure consists of two independent 7 SL RNA-like folding units. *J. Biol. Chem.* **266:** 8675-8678.

Sippel, A.E., Kurtz, D.T., Morris, H.P., and Feigelson, P. (1976). Comparison of *in vivo* translation rates and messenger RNA levels of α_{2u}-globulin in rat liver and Morris hepatoma 5123D. *Cancer Res.* **36:** 3588-3593.

Slootweg, M.C., deGroot, R.P., Herrmann, E.M.P., Koorneef, I., Kruijer, W., and Kramer, Y.M. (1991). Growth hormone induces expression of c-jun and jun B oncogenes and employs a protein kinase C signal transduction pathway for the induction of c-fos oncogene expression. *J. Mol. Endocrin.* **6:** 179-188.

Slootweg, M.C., van Genesen, S.T., Otte, A.P., Duursma, S.A., and Kruijer, W. (1990). Activation of mouse osteoblast growth hormone receptor: c-fos oncogene expression independent of phosphoinositide breakdown and cyclic AMP. *J. Mol. Endocrin.* **4:** 265-274.

Smith, M.M. (1984). The organization of the yeast histone genes. *In* "Histone Genes. Structure, Organization and Regulation". pp. 3-33, John Wiley & Sons, New York.

Smoller, D., Petrov, D, and Hartl, D. (1991). Characterization of bacteriophage P1 library containing inserts of Drosophila DNA of 75-100 kilobase pairs. *Chromosoma* **100:** 487-494.

Soares, V.C., Gubits, R.M., Feigelson, P., and Costantini, F. (1987). Tissue-specific and hormonally regulated expression of a rat α_{2u}-globulin gene in transgenic mice. *Mol. Cell. Biol.* **7:** 3749-3758.

Southard-Smith, M., and MacDonald, R.J. (1993). Isolating reiterated genes from a rat genomic library constructed with the bacteriophage P1 system. *DuPont Biotech Update* **summer:** 36-39.

Southard-Smith, M., Pierce, J.C., and MacDonald, R.J. (1994). Physical mapping of the rat-tissue kallikrein family in 2 gene clusters by analysis of P1-bacteriophage clones. *Genomics* **22:** 404-417.

Southern, E.M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* **98:** 503-517.

Southern, E.M. (1992). Genome mapping: cDNA approaches. *Curr. Opin. Genet. Dev.* **2:** 412-416.

Stallings, R.L., Ford, A.F., Nelson, D., Torney, D.C., Hildebrand, C.E., and Moyzis, R.K. (1991). Evolution and distribution of $(GT)_n$ repetitive sequences in mammalian genomes. *Genomics* **10:** 807-815.

Stein, G.S., Sierra, F., Stein, J.L., Plumb, M., Marashi, F., Carozzi, N., Prokopp, K., and Baumbach, L. (1984). Organization and expression of the human histone genes. *In* "Histone Genes. Structure, Organization and Regulation". pp. 397-455, John Wiley & Sons, New York.

Sternberg, N. (1990). A bacteriophage P1 cloning system for the isolation, amplification and recovery of DNA fragments as large as 100 kbp. *Proc. Natl. Acad. Sci. USA* **87:** 103-107.

Sternberg, N. (1994). The P1 cloning system-past and future. *Mamm. Genome* **5:** 397-404.

Sternberg, N., Ruether, J., and DeRiel, K. (1990). Generation of a 50,000-member human DNA library with an average insert size of 75-100 kbp in a bacteriophage P1 cloning vector. *New Biol.* **2:** 151-162.

Swenberg, J.A. (1993). α_{2u}-globulin nephropathy: Review of the cellular and molecular mechanisms involved and their implications for human risk assessment. *Environ. Health Perspect. Suppl.* **101:** 39-44.

Swenberg, J.A., Short, B., Borghoff, S., Strasser, J., and Charbonneau, M. (1989). The comparative pathobiology of α_{2u}-globulin nephropathy. *Toxicol. Appl. Pharmacol.* **97:** 35-46.

Szpirer, C., Riviere, M., Genet, M., Dreze, P., Islam, M.Q., and Levan, G. (1990). Assignment of 12 loci to rat chromosome 5: Evidence that this chromosome is homologous to mouse chromosome 4 and to human chromosomes 9 and 1 (1p arm). *Genomics* **6:** 679-684.

Szpirer, C., Szpirer, J., Islam, M.Q., and Levan, G. (1988). The rat gene map. *Curr. Top. Microbiol. Immunol.* **137:** 33-38.

Takano, K., Hall, K., Fryklund, , L., Holmgren, A., Sievertsson, H., and Uthne, K. (1975). The binding of insulin and somatomedin A to human placental membrane. *Acta Endocrin.* **80:** 14-31.

Tartof, K.D., and Hobbs, C.A. (1988). New cloning vectors and techniques for easy and rapid restriction mapping. *Gene* **67:** 169-182.

Tautz, D., and Renz, M. (1984). simple sequences are ubiquitous repetitive components of eukaryotic genomes. *Nucl. Acids Res.* **12:** 4127-4138.

Tautz, D., Trick, M., and Dover, G. (1986). Cryptic simplicity in DNA is a major source of genetic variation. *Nature* **322:** 652-656.

Thomashow, L.S., Milhausen, M., Rutter, W.J., and Agabian, N. (1983). Tubulin genes are tandemly linked and clustered in the genome of *Trypanosoma brucei*. *Cell* **32:** 35-43.

Tocharoentanaphol, C., Cremer, M., Schrock, E., Blonden, L., Kilian, K., Cremer, T., and Ried, T. (1994). Multicolor fluorescence in situ hybridization on metaphase chromosomes and interphase Halo-preparations using cosmid and YAC clones for the simultaneous high resolution mapping of deletions in the dystrophin gene. *Hum. Genet.* **93:** 229-235.

Topp, W.C. (1981). Normal rat cell lines deficient in nuclear thymidine kinase. *Virology* **113:** 408-411.

Trask, B., Pinkel, D., and van den Engh, G. (1989). The proximity of DNA sequences in interphase cell nuclei is correlated to genomic distance and permits ordering of cosmids spanning 250 kilobase pairs. *Genomics* **5:** 710-717.

Trask, B.J., Massa, H., Kenwrick, S., and Gitschier, J. (1991). Mapping of human chromosome Xq28 by two color fluorescence in situ hybridization of DNA sequences to interphase cell nuclei. *Am. J. Hum. Genet.* **48:** 1-15.

Traver, C.N., Klapholz, S., Hyman, R.W., and Davis, R.W. (1989). Rapid screening of a human genomic library in yeast artificial chromosomes for single-copy sequences. *Proc. Natl. Acad. Sci. USA* **86:** 5898-5902.

Trowsdale, J., Ragoussis, J., and Campbell, R.D. (1991). Map of the human MHC. *Immunol. Today* **12:** 443-446.

Tschudi, C., Pirrotta, V., and Junakovic, N. (1982). Rearrangements of the 5 S RNA gene cluster of *Drosophila melanogaster* associated with the insertion of a B104 element. *EMBO J.* **1:** 977-985.

Tumlinson, J.H., Yonce, C.E., Doolittle, R.E., Heath, R.R., Gentry, C.R., and Mitchell, E.R. (19774). Sex pheromones and reproductive isolation of the lesser peachtree borer and the peachtree borer. Science 185: 614-616.

Unterman, R.D., Lynch, K.R., Nakhasi, H.L., Dolan, K.P., Hamilton, J.W., Cohn, D.V., and Feigelson, P. (1981). Cloning and sequence of several alpha-2u-globulin cDNAs. *Proc. Natl. Acad. Sci. USA* **78:** 3478-3482.

Van Arsdell, S.W., and Weiner, A.M. (1984). Human genes for U2 small nuclear RNA are tandemly repeated. *Mol. Cell. Biol.* **4:** 492-499.

Van Cong, N., Vaisse, C., Gross, M.-S., Slim, R., Milgrom, E., and Bernheim, A. (1991). The human placental protein 14 (PP14) gene is localized on chromosome 9q34. *Hum. Genet.* **86:** 515-518.

Van den Berg, M.J., and Ziegelberger, G. (1991). On the function of the pheromone binding protein in the olfactory hairs of *Antheraea polyphemus*. *J. Insect Physiol.* **37:** 79-85.

Van den Engh, G., Sachs, R., and Trask, B.J. (1992). Estimating genomic distance from DNA sequence location in cell nuclei by a random walk model. *Science* **257:** 1410-1412.

Van Dijck, P., and Verhoevan, G. (1992). Interaction of estrogen receptor complexes with the promoter region of genes that are negatively regulated by estrogens: the alpha-2u-globulins. *Biochem. Biophys. Res. Comm.* **182:** 174-181.

Van Dijck, P., De Vos, P., Winderickx, J., and Verhoeven, G. (1993a). Multiple binding sites for nuclear factors in the 5' upstream region of two alpha-2u-globulin genes: Implications for hormone-regulated and tissue-specific control. *J. Steroid Biochem. Mol. Biol.* **45:** 353-366.

Van Dijck, P., Lemaigre, F.P., Winderickx, J., Rousseau, G.G., and Verhoeven, G, (1987). Glucocorticoid receptor binding to defined regions of α_{2u}-globulin genes. *Biochem. Biophys. Res. Comm.* **149:** 1099-1105.

Van Dijck, P., Schoonjans, K., Sassone-Corsi, P., Auwerx, J., and Verhoevan, G. (1993b). A Fos-Jun element in the first intron of an alpha-2u-globulin gene. *Mol. Cell. Biochem.* **125:** 127-136.

Van Dongen, W.M.A.M., Moorman, A.F.M., and Destree, O.H.J. (1984). DNA organization and expression of *Xenopus* histone genes. *In* "Histone Genes. Structure, Organization and Regulation". pp. 199-224, John Wiley & Sons, New York.

Van Wyk, J.J., Underwood, L.F., Baseman, J.B., Hintz, R.L., Clemmons, D.R., and Marshall, R.N. (1975). Explorations of the insulin like and growth-promoting properties of somatomedin by membrane receptor assays. *Adv. Metab. Disord.* **8:** 127-150.

Vandenberg, J.G. (1969). Male odor accelerates female sexual maturation in mice. *Endocrinology* **84:** 658-660.

Vandenberg, J.G., Whitsett, J.M., and Lombardi, J.R. (1975). Partial isolation of a pheromone accelerating puberty in female mice. *J. Reprod. Fertil.* **43:** 515-523.

Virgilio, L., Isobe, M., Narducci, M.G., Carotenuto, P., Camerini, B., Kurosawa, N., Rushdi, A., Croce, C., and Russo, G. (1993). *Proc. Natl. Acad. Sci. USA* **90:** 9275-9279.

Vogt, R.G., Prestwich, G.D., and Riddiford, L.M. (1988). Sex pheromone receptor proteins. Visualization using a radiolabeled photoaffinity analog. *J. Biol. Chem.* **263:** 3952-3959.

Vokel-Meiman, K., and Roeder, G.S. (1990). Gene conversion tracts stimulated by HOT1-promoted transcription are long and continuous. *Genetics* **126:** 851-867.

Voliva, C.F., Jahn, C.L., Comer, M.B., Hutchinson, C.A., and Edgell, M.H. (1983). The L1Md long interspersed repeat family in the mouse: Almost all examples are truncated at one end. *Nucl. Acids Res.* **11:** 8847-8859.

Wada, M., Little, R.D., Abid, F., Porta, G., Labella, T., Cooper, T., Della Valle, G., D'Urso, M., and Schlessinger, D. (1990): Human Xq24-28: approaches to mapping with yeast artificial chromosome. *Am. J. Hum. Genet.* **46:** 95-106.

Waldman, A.S., and Liskay, R.M. (1987). Differential effects of base-pair mismatch on intrachromosomal versus extrachromosomal recombination in mouse cells. *Proc. Natl. Acad. Sci. USA* **84:** 5340-5344.

Walsh, J.B. (1995). How often do duplicated genes evolve new functions? *Genetics* **139:** 421-428.

Wang, K.S. (1996). Hormonal regulation of rat α_{2u}-globulin genes. Ph. D. thesis. University of Alberta, Edmonton.

Ward, E.R., and Jen, G.C. (1990). Isolation of single-copy-sequence clones from a yeast artificial chromosome library of randomly sheared *Arabidopsis thaliana* DNA. *Plant Mol. Biol.* **14:** 561-568.

Watanabe, N., and Ohshima, Y. (1988). Three types of rat U1 small nuclear RNA genes with different flanking sequences are induced to express *in vivo*. *Eur J. Biochem.* **174:** 125-132.

Weber, J.L. (1990). Informativeness of human $(dC\text{-}dA)_n$ $(dG\text{-}dT)_n$ polymorphisms. *Genomics* **7:** 524-530.

Weissenbach, J., Gyapay, G., Dib, C., Vignal, A., Morissette, J., Millasseau, P., Vaysseix, G., and Lathrop, M. (1992). A second generation linkage map of the human genome. *Nature* **359:** 794-801.

Wheeler, C.J., Maloney, D., and Fogel, S. (1990). Microconversion between murine H-2 genes integrated into yeast. *Nature* **347:** 192-194.

White, D.M., Mikol, D.D., Espinosa, R., Weimer, B., LeBeau, M.M., and Stefansson, K. (1992). Structure and chromosomal localization of the human gene for a brain form of prostaglandin D-2 synthase. *J. Biol. Chem.* **267:** 23202-23208.

Widman, L.E., and Chasin, L.A. (1982). Multihormonal induction of α_{2u} globulin in an established rat hepatoma cell line. *J. Cell. Physiol.* **112:** 316-326.

Wiegant, J., Kalle, W., Mullenders, L., Brookes, S., Hoovers, J.M.N., Dauwerse, J.G., Van Ommen, G.J.B., and Raap, A.K. (1992). High-resolution *in situ* hybridization using DNA halo preparations. *Hum. Mol. Genet.* **1:** 587-591.

Willard, H.F., and Waye, J.S. (1987). Chromosome-specific subsets of human alpha satellite DNA: Analysis of sequence divergence within and between chromosomal subsets and evidence of an ancestral pentameric repeat. *J. Mol. Evol.* **25:** 207-214.

Winderickx, J., Van Dijck, P., Dirckx, L., Volckaert, G., Rambauts, W., Heyns, W., and Verhoeven, G. (1987). Comparison of the 5' upstream putative regulatory sequences of three members of the α 2u globulin gene family. *Eur. J. Biochem.* **165:** 521-529.

Woodcock, D.M., Crowther, P.J., Doherty, J., Jefferson, S., DeCruz, E., Noyer-Weidner, M., Smith, S.S., Michael, M.Z., and Graham, M.W. (1989). Quantitative evaluation of Escherichia coli host strains for tolerance to cytosine methylation in plasmid and phage recombinants. *Nucl. Acids Res.* **17:** 346-378.

Xiong, Y., Sakaguchi, B., and Eickbush, T.H. (1988). Gene conversions can generate sequence variants in the late chorion multigene families of *Bombyx mori*. *Genetics* **120:** 221-231.

Yamamoto, M., Gao, F., Furuichi, M., Ichiyoshi, Y., and Endo, H. (1989). A LINE 1 sequence interupts the rat α2u globulin gene. *Biochim. Biophys. Acta* **1008:** 322-328.

Yanai, J., and McClearn, G.E. (1973). assortive mating in mice 3 Genetic determination of female mating preference. *Behav. Genet.* **3:** 75-84.

Yoon, J.B., Berry, S.A., Seelig, S., and Towle, H.C. (1990). An inducible nuclear factor binds to a growth hormone-regulated gene. *J. Biol. Chem.* **265:** 19947-19954.

Yoshida, M.C. (1978). Rat gene mapping by rat-mouse somatic cell hybridization and a comparative Q-banding analysis between rat and mouse chromosomes. *Cytogenet. Cell Genet.* **22:** 606-609.

Zakian, V.A. (1989). Structure and function of telomeres. *Ann. Rev. Genet.* **23:** 579-604.

Zeller, R., Carri, M.T., Mattaj, I.W., and DeRobertis, E.M. (1984). *Xenopus laevis* U1 snRNA genes: Characterization of transcriptionally active genes reveals major and minor repeated gene families. *EMBO J.* **3:** 1075-1081.

Zeng, C., Spielman, A.I., Vowels, B.R., Leydens, J.J., Biemann, K., and Preti, G. (1996). A human axillary odorant is carried by apolipoprotein D. *Proc. Natl. Acad. Sci. USA* **93:** 6626-6630.

APPENDIX

Table A.1. Physical Analysis of Fiber FISH Images of p2860

Molecule	Scale	Left End-Gene 1 Interval	Gene 1 Interval	Gene 1-Gene 2 Interval	Gene 2 Interval	Gene 2-Gene 3 Interval	Gene 3 Interval	Gene 3-Right End Interval	Total Interval Values	Stretch (µm/kb)
	µm	0.62	0.62	7.75	1.09	6.51	1.40	5.27	23.26	
1	%	2.67	2.67	33.32	4.69	27.98	6.01	22.66	100.00	0.256
	kb	2.4	2.4	30.3	4.3	25.5	5.5	20.6	91.0	
	µm	0.47	1.40	9.61	1.24	8.37	1.40	3.57	26.06	
2	%	1.80	5.37	36.88	4.76	32.12	5.37	13.70	100.00	0.286
	kb	1.6	4.9	33.6	4.3	29.2	4.9	12.5	91.0	
	µm	1.24	1.24	5.43	1.86	6.51	2.02	6.36	24.66	
3	%	5.03	5.03	22.02	7.54	26.40	8.19	25.79	100.00	0.271
	kb	4.6	4.6	20.0	6.9	24.0	7.4	23.5	91.0	
	µm	0.93	0.78	3.72	0.78	5.12	0.78	1.71	13.82	
4	%	6.73	5.64	26.92	5.64	37.05	5.64	12.38	100.00	0.152
	kb	6.1	5.1	24.5	5.1	33.7	5.1	11.4	91.0	
	µm	0.93	0.78	8.68	1.86	4.34	1.40	6.36	24.35	
5	%	3.82	3.20	35.65	7.64	17.82	5.75	26.12	100.00	0.268
	kb	3.5	2.9	32.4	7.0	16.2	5.2	23.8	91.0	
	µm	0.47	2.48	11.63	1.55	13.95	1.40	3.88	35.36	
6	%	1.33	7.01	32.89	4.38	39.45	3.96	10.98	100.00	0.389
	kb	1.2	6.4	29.9	4.0	35.9	3.6	10.0	91.0	
	µm	0.47	1.24	5.27	1.86	10.54	1.24	3.72	24.34	
7	%	1.94	5.09	21.65	7.64	43.30	5.09	15.28	100.00	0.267
	kb	1.8	4.6	19.7	6.9	39.4	4.6	14.0	91.0	

Table A.1. continued Physical Analysis of Fiber FISH Images of p2860

Molecule	Scale	Left End-Gene 1 Interval	Gene 1 Interval	Gene 1-Gene 2 Interval	Gene 2 Interval	Gene 2-Gene 3 Interval	Gene 3 Interval	Gene 3-Right End Interval	Total Interval Values	Stretch (μm/kb)
	μm	0.47	1.71	7.13	1.40	7.29	1.55	2.33	21.88	
8	%	2.15	7.82	32.58	6.40	33.32	7.08	10.65	100.00	0.240
	kb	2.0	7.1	29.6	5.8	30.4	6.4	9.7	91.0	
	μm	0.78	1.86	4.34	1.71	10.85	1.55	2.95	24.04	
9	%	3.24	7.74	18.05	7.11	45.13	6.45	12.28	100.00	0.264
	kb	2.9	7.0	16.4	6.5	41.1	5.9	11.2	91.0	
Average	kb	2.9±1.1	5.0±1.7	26.3±9.5	5.6±2.0	30.6±11.1	5.4±1.9	15.2±4.1	91.0	0.266
Expected	kb	4.0	4.5	21.0	4.5	32.5	4.5	20.0	91.0	0.340

Measurements for the specified intervals on the 9 molecules of p2860 shown schematically in Figure 4.4 are presented. The average values in kb are given, with their associated standard error, as are the expected values, based on restriction endonuclease analysis. The value given for the the expected degree of DNA stretching is that predicted for relaxed duplex DNA.

Table A.2. Physical Analysis of Fiber FISH Images of p2861

Molecule	Scale	Left End-Gene 1 Interval	Gene 1 Interval	Gene 1-Right End Interval	Total Interval Values	Stretch (µm/kb)
	µm	4.34	0.93	12.40	17.67	
1	%	24.56	5.26	70.18	100.00	0.188
	kb	23.1	4.9	66.0	94.0	
	µm	6.82	1.24	22.33	30.39	
2	%	22.44	4.08	73.48	100.00	0.323
	kb	21.1	3.8	69.1	94.0	
	µm	7.44	1.55	13.33	22.32	
3	%	33.34	6.94	59.72	100.00	0.237
	kb	31.3	6.5	56.2	94.0	
	µm	6.20	1.24	15.81	23.25	
4	%	26.67	5.33	68.00	100.00	0.247
	kb	25.1	5.0	63.9	94.0	
	µm	7.44	2.48	20.47	30.39	
5	%	24.48	8.16	67.36	100.00	0.323
	kb	23.0	7.7	63.3	94.0	
	µm	5.89	1.24	20.16	27.29	
6	%	21.58	4.55	73.87	100.00	0.290
	kb	20.3	4.3	69.4	94.0	
	µm	3.72	1.55	12.71	17.98	
7	%	20.69	8.62	70.69	100.00	0.191
	kb	19.5	8.1	66.4	94.0	

Table A.2. continued Physical Analysis of Fiber FISH Images of p2861

Molecule	Scale	Left End-Gene 1 Interval	Gene 1 Interval	Gene 1-Right End Interval	Total Interval Values	Stretch (µm/kb)
	µm	7.13	1.40	25.11	33.64	
8	%	21.20	4.16	74.64	100.00	0.358
	kb	19.9	3.9	70.2	94.0	
	µm	6.82	1.24	13.33	21.39	
9	%	31.88	5.80	62.32	100.00	0.228
	kb	30.0	5.5	58.5	94.0	
	µm	8.68	1.09	14.57	24.34	
10	%	35.66	4.48	59.86	100.00	0.259
	kb	33.5	4.2	56.3	94.0	
Average	kb	24.7±8.4	5.4±1.8	63.9±21.4	94.0	0.264
Expected	kb	23	4.5	66.5	94.0	0.340

Measurements for the specified intervals on the 10 molecules of p2861 shown schematically in Figure 4.5 are presented. The average values in kb are given, with their associated standard error, as are the expected values, based on restriction endonuclease analysis. The value given for the the expected degree of DNA stretching is that predicted for relaxed duplex DNA.

Table A.3. Physical Analysis of Fiber FISH Images of p2862

Molecule	Scale	Left End-Gene 1 Interval	Gene 1 Interval	Gene 1-Gene 2 Interval	Gene 2 Interval	Gene 2-Right End Interval	Total Interval Values	Stretch (µm/kb)
	µm	7.59	1.86	5.89	2.17	13.64	31.15	
1	%	24.36	5.97	18.91	6.97	43.78	100.00	0.302
	kb	25.1	6.1	19.5	7.2	45.1	103.0	
	µm	8.06	1.40	4.19	1.86	9.92	25.43	
2	%	31.70	5.51	16.48	7.31	39.00	100.00	0.246
	kb	32.6	5.7	17.0	7.5	40.2	103.0	
	µm	6.66	1.24	4.65	0.93	12.70	26.18	
3	%	25.44	4.74	17.76	3.55	48.51	100.00	0.254
	kb	26.2	4.9	18.3	3.7	49.9	103.0	
	µm	6.04	1.08	3.88	1.09	11.78	23.78	
4	%	25.30	4.53	16.25	4.57	49.35	100.00	0.232
	kb	26.1	4.7	16.7	4.7	50.8	103.0	
	µm	5.43	1.55	5.89	1.40	11.78	26.05	
5	%	20.85	5.95	22.61	5.37	45.22	100.00	0.253
	kb	21.5	6.1	23.3	5.5	46.6	103.0	
	µm	5.73	1.24	2.95	0.93	7.44	18.29	
6	%	31.33	6.78	16.13	5.08	40.68	100.00	0.178
	kb	32.3	7.0	16.6	5.2	41.9	103.0	
	µm	7.44	1.86	5.58	1.71	14.61	31.20	
7	%	23.85	5.96	17.88	5.48	46.83	100.00	0.303
	kb	24.6	6.1	18.4	5.7	48.2	103.0	

Table A.3. continued Physical Analysis of Fiber FISH Images of p2862

Molecule	Scale	Left End-Gene 1 Interval	Gene 1 Interval	Gene 1-Gene 2 Interval	Gene 2 Interval	Gene 2-Right End Interval	Total Interval Values	Stretch (μm/kb)
	μm	7.75	1.24	3.57	1.09	8.68	22.33	
8	%	34.71	5.55	15.99	4.88	38.87	100.00	0.218
	kb	35.8	5.7	16.5	5.0	40.0	103.0	
	μm	6.20	0.93	4.65	1.55	7.75	21.08	
9	%	29.41	4.41	22.06	7.35	36.77	100.00	0.205
	kb	30.3	4.5	22.7	7.6	37.9	103.0	
	μm	4.96	0.78	4.19	0.78	12.09	22.80	
10	%	21.75	3.42	18.38	3.42	53.03	100.00	0.221
	kb	22.5	3.5	18.9	3.5	54.6	103.0	
	μm	5.43	0.93	3.72	1.09	8.68	19.85	
11	%	27.36	4.69	18.74	5.48	43.73	100.00	0.193
	kb	28.2	4.8	19.3	5.6	45.1	103.0	
Average	kb	27.7±8.9	5.4±1.7	18.8±6.0	5.6±1.8	45.5±14.4	103.0	0.237
Expected	kb	35	4.5	13	4.5	46	103.0	0.340

Measurements for the specified intervals on the 11 molecules of p2862 shown schematically in Figure 4.6 are presented. The average values in kb are given, with their associated standard error, as are the expected values, based on restriction endonuclease analysis. The value given for the the expected degree of DNA stretching is that predicted for relaxed duplex DNA.